KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

2ND ACTUARIAL AND FINANCIAL MATHEMATICS DAY

6 februari 2004

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

CONTACTFORUM

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

2ND ACTUARIAL AND FINANCIAL MATHEMATICS DAY

6 februari 2004

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

CONTACTFORUM

Handelingen van het contactforum "2nd Actuarial and Financial Mathematics Day" (6 februari 2004, 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/2004/0455/20

Printed by *Universa Press, 9230 Wetteren, Belgium*

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

Contactforum: 2nd Actuarial and Financial Mathematics Day

PREFACE

These proceedings contain the contributions of the 2nd Actuarial and Financial Mathematics day, held in Brussels on February 6, 2004. In a few words the items that have been treated range from pricing derivatives and portfolio selection over interest rate modelling to risk theory and non-life insurance. It just shows that "Actuarial and Financial Mathematics" covers a large domain of interest.

For the second edition of this day we are very grateful to the Royal Flemish Academy of Sciences and Art for their sponsoring and hospitality. They created the opportunity to meet in a superb accommodation, breathing an atmosphere of wisdom.

The intention of this meeting is to gather junior researchers and postdocs, active in either financial or more actuarial sciences and to give them a forum to show their progress in a domain which covers a broad spectrum and nevertheless is still a beautiful, attractive and stimulating area. Besides the two invited talks, eight contributed talks were given.
This meeting is an occasion to maintain the bridge – to build just means there is none – between practice and theory. So we were very glad to welcome a lot of participants from banks and insurance companies. Also one of the speakers was a practitioner from insurance.

The success of this contactforum is a great stimulation for the organisers to continue with this yearly initiative.

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

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

Contactforum: 2nd Actuarial and Financial Mathematics Day

TABLE OF CONTENTS

Invited talks

Contributed talks

INVITED TALKS

THE VALUATION OF ASIAN OPTIONS FOR MARKET MODELS OF EXPONENTIAL LÉVY TYPE

Hansjörg Albrecher

Graz University of Technology, Austria and K.U.Leuven, Belgium
Email:`albrecher@TUGraz.at`

Abstract

We give a brief survey on some recent developments in pricing and hedging of European-style arithmetic average options given that the underlying asset price process is of exponential Lévy type.

1. INTRODUCTION

During the last decade it has been realized that the strong assumptions of the classical Black-Scholes model for the stochastic behavior over time of stock prices and indices are usually not fulfilled in practical applications. Among the major deficiencies of the Black-Scholes model are the normality assumption of log returns across all time scales and the assumption of a non-stochastic volatility (see e.g. [36]). In this survey, we will consider finance market models of exponential Lévy type which are able to capture the empirically observed distributional behavior of log returns and thus overcome some imperfections of the Black-Scholes model. In particular, we will discuss the pricing and hedging of Asian options under these market models. It turns out that by exploiting the independent and stationary increments property of Lévy processes, one can derive quick and rather accurate approximations of Asian option prices for arbitrary risk-neutral pricing measures (Section 4). In Section 5 a simple static super-hedging strategy for the payoff of Asian options in terms of a portfolio of European options is discussed. Its performance can be optimized by utilizing comonotonicity theory. This hedging strategy can even be applied to market models including stochastic volatility.

2. THE EMPIRICAL BEHAVIOR OF LOG RETURNS

One of the crucial assumptions of the Black-Scholes model is that log returns across all time scales follow a normal distribution. However, this is clearly unrealistic in practice, in particular for short

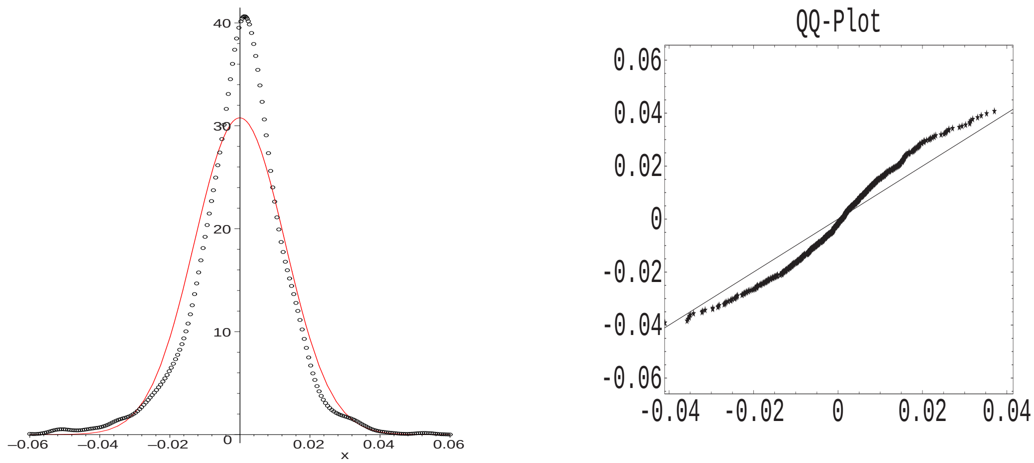


Figure 1: Kernel density estimator (dotted line) and maximum-likelihood fit of a normal distribution (solid line) for daily ATX log returns (Jan 97 - June 00) and corresponding QQ-plot.

time horizons. As an illustrative example, Figure 1 shows a kernel density estimator of daily log-returns of the Austrian stock index (ATX) based on data over a time span of more than three years in comparison to a maximum-likelihood fit of a normal distribution as proposed by the Black-Scholes model. Clearly the normal distribution does not reflect the empirical distribution, lacking mass in the center and in the tails. Various alternatives have been proposed for fitting log returns, among them the generalized hyperbolic (GH) distribution (cf. [21]), the Meixner distribution (cf. [35]) and the CGMY distribution [13, 14].

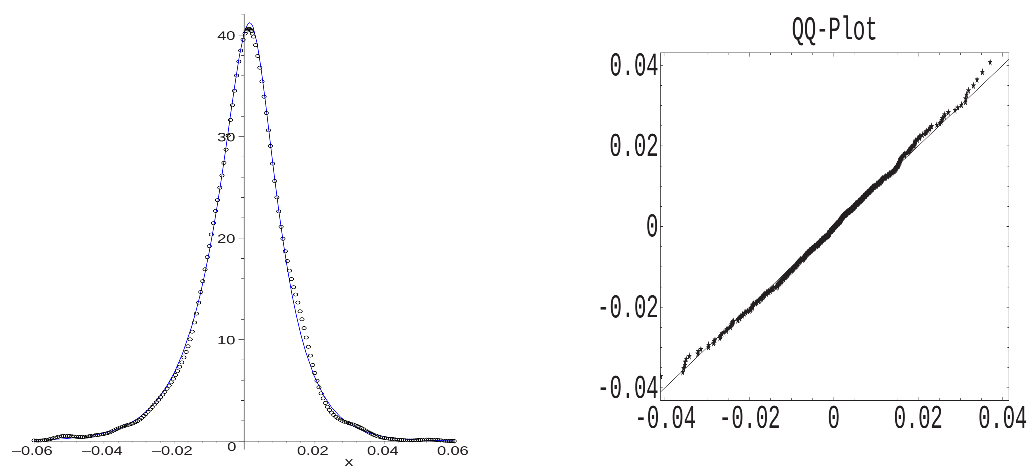


Figure 2: Kernel density estimator (dotted line) and maximum-likelihood fit of a GH distribution (solid line) for daily ATX log returns (Jan 97 - June 00) and corresponding QQ-plot.

Figure 2 shows a maximum-likelihood fit of the GH distribution to the same data set and it can be observed that the GH distribution is able to capture both the behavior at the center and in the (semi-heavy) tails quite well. The GH distribution (originally introduced by Barndorff-Nielsen [5]) has five parameters and contains the normal, the normal inverse Gaussian and the variance gamma distribution as a special case (for a detailed discussion see [8]). For further investigations on the suitability of these distributions for fitting financial data we refer to [6, 24, 29, 31, 34]. GH, Meixner and CGMY distributions all are infinitely divisible and thus generate a Lévy process as described in the next section.

3. THE EXPONENTIAL LÉVY MODEL

Any infinitely divisible distribution X generates a Lévy process $(Z_t)_{t\geq 0}$, i.e. a stochastic process with stationary and independent increments, $Z_0 = 0$ a.s. and $Z_1 \sim^d X$. It is always possible to choose a càdlàg version of the Lévy process. According to the construction, increments of length 1 have distribution X, but in general none of the increments of length different from 1 has a distribution of the same class. Exceptions are the cases where X is a normal inverse Gaussian, a variance gamma, a Meixner or a CGMY distribution; due to their respective convolution properties, each increment $Z_t - Z_s \, (t > s \geq 0)$ is then of the same class with new parameters depending on $t - s$, which makes these distributions natural and particularly attractive candidates for fitting the marginal distributions.

An exponential Lévy model for the price process $(S_t)_{t\geq 0}$ of an asset (a stock or an index) is now defined by

$$S_t = S_0 \, \exp(Z_t). \tag{1}$$

Thus the Brownian motion with drift of the Black-Scholes model is replaced by a Lévy process. Indeed, the distribution of log returns over time t is now given by the distribution of Z_t. Implications of the model choice (1) on the dynamics of the asset price process are e.g. discussed in [20]. There is some empirical evidence that the Lévy measure of realistic Lévy models does not contain a Brownian ”diffusion” component, so that the price process $(S_t)_{t\geq 0}$ is purely discontinuous (with finitely or infinitely many jumps in every finite interval, see e.g. [13])). The model (1) assumes a constant volatility, but the volatility smile effect of the Black-Scholes model is considerably reduced (cf. [24]). Time-consistency of Lévy models was investigated in [22]. For an up-to-date survey on exponential Lévy models we refer to [20, 36].

The market model (1) is in general incomplete (cf. Cherny [19]), and there exist infinitely many equivalent martingale measures $\mathbb{Q}$ so that in order to price derivative securities one has to choose one particular candidate. One mathematically tractable choice is the so-called Esscher equivalent measure, essentially obtained by exponential tilting of the original measure. It was first introduced to mathematical finance by Madan and Milne [29]; see also Gerber and Shiu [25]. Let f_t denote the density of the marginal distribution Z_t, then the Esscher transform of f_t is defined by

$$f_t(x;\theta) = \frac{\mathrm{e}^{\theta x} \, f_t(x)}{\int_{-\infty}^{\infty} \mathrm{e}^{\theta y} \, f_t(y) \, \mathrm{d}y}$$

with $\theta \in \mathbb{R}$. One can now define another Lévy process $(Z_t^\theta)_{t\geq 0}$ such that its one-dimensional marginal distributions are the Esscher transforms of the corresponding marginals of $(Z_t)_{t\geq 0}$ (for details see Raible [32]) and the parameter θ is chosen in such a way, that the discounted stock price process $(\mathrm{e}^{-r\,t} \, S_t^\theta)_{t\geq 0}$ is a $\mathbb{Q}$-martingale (where r denotes the risk-free interest rate). It turns out that for normal inverse Gaussian and variance gamma Lévy processes the switch to the Esscher measure just amounts to a shift in the parameters (cf. [2, 3]), which makes the analysis particularly tractable. There have been attempts to justify this particular choice for $\mathbb{Q}$ both within utility and

equilibrium theory; however, the topic is still controversial (cf. [11, 18, 26]).
Another natural approach is to shift the drift of the Lévy process in such a way that a risk-neutral framework is obtained. However, in this case the resulting risk-neutral measure is in general not equivalent to the physical measure. One way to circumvent this problem is to start out immediately with a risk-neutral model for S_t defined by

$$S_t = S_0 \frac{\exp(r\,t)}{E[\exp(Z_t)]} \exp(Z_t)$$

and then calibrating the parameters from current option prices observed in the market rather than from historical log-returns. This approach is quite common in practice, see e.g. [36].
Note that the techniques presented in Sections 4 and 5 are applicable for any risk-neutral pricing measure $\mathbb{Q}$ and thus the choice of $\mathbb{Q}$ is not discussed any further.

4. PRICING OF ASIAN OPTIONS

Let us now consider the price of a European-style arithmetic average call option at time t under exponential Lévy models given by

$$\mathrm{AA}_t = \frac{\mathrm{e}^{-r(T-t)}}{n}\, \mathbb{E}^{\mathbb{Q}} \left[\left(\sum_{k=0}^{n-1} S_{T-k} - nK \right)^+ \,\middle|\, \mathcal{F}_t \right], \qquad (2)$$

where n is the number of averaging days, K the strike price, T the time to expiration, $\mathcal{F}_t$ the information available at time t and $\mathbb{Q}$ any risk-neutral pricing measure. For convenience we will restrict ourselves to the case $t = 0$ and $n = T$, so that the averaging starts at time 1 (the other cases can be handled in a completely analogous way).

The main difficulty in evaluating (2) is to determine the distribution of the dependent sum $\sum S_k$, for which in general no explicit analytical expression is available. There are several approaches to the problem: one can use Monte Carlo simulation techniques to obtain numerical estimates of the price, which can be achieved by adapting procedures developed for the Black-Scholes case (see e.g. [9, 10, 33, 37]). Recently Večeř and Xu [40] developed a partial integro-differential equation approach that is applicable for exponential Lévy models, which transforms the problem into finding numerical solution of these equations. Both approaches are rather time consuming. For an approach based on Fast Fourier Transforms, see [7, 17]. Another alternative is to use approximations of the distribution of the average, which sometimes leads to closed-form expressions for the price approximation. In the sequel we will discuss an adaption of such approximation techniques developed for the Black-Scholes case ([28, 39, 41]) to our exponential Lévy setting.
The basic idea is to determine moments of the dependent sum in (2) and then replace it by a more tractable distribution with identical first moments. Due to the independence and stationarity of increments of Lévy processes, one can derive a simple algorithm to derive the mth moment of the dependent sum $A_n := \sum_{k=1}^{n} S_k$:

Let us define

$$R_i = \frac{S_i}{S_{i-1}}, \quad i = 1, \ldots, n$$

and

$$L_n = 1$$
$$L_{i-1} = 1 + R_i L_i, \quad i = 2, \dots, n.$$

Then we have

$$\sum_{k=1}^{n} S_k = S_0(R_1 + R_1R_2 + \cdots + R_1R_2 \dots R_n) = S_0R_1L_1.$$

Thus it remains to determine $\mathbb{E}^{\mathbb{Q}}[(R_1L_1)^m] = \mathbb{E}^{\mathbb{Q}}[R_1^m]\mathbb{E}^{\mathbb{Q}}[L_1^m]$ (the last equality follows from the independence of the increments). But

$$\mathbb{E}^{\mathbb{Q}}[R_i^k] = \mathbb{E}^{\mathbb{Q}}[\exp(k\, Z_1)] = \mathbb{E}^{\mathbb{Q}}[\exp(k\, X)], \tag{3}$$

so that one just has to evaluate the risk-neutral moment generating function of X at k, given it exists (recall that X is the generating distribution of the Lévy process). Furthermore we have

$$\mathbb{E}^{\mathbb{Q}}[L_{i-1}^m] = \mathbb{E}^{\mathbb{Q}}[(1 + L_iR_i)^m] = \sum_{k=0}^{m} \binom{m}{k} \mathbb{E}^{\mathbb{Q}}[L_i^k]\mathbb{E}^{\mathbb{Q}}[R_i^k]. \tag{4}$$

Starting with $\mathbb{E}^{\mathbb{Q}}[L_n^k] = 1 \;\forall\; k \in \{0, \dots, m\}$, one can then apply recursion (4) together with (3) to obtain $\mathbb{E}^{\mathbb{Q}}[L_1^m]$ and subsequently $\mathbb{E}^{\mathbb{Q}}[(A_n)^m] = \mathbb{E}^{\mathbb{Q}}[R_1^m]\mathbb{E}^{\mathbb{Q}}[L_1^m]$.

These moments can now be used to approximate $A_n = \sum_{k=1}^{n} S_k$ by another more tractable distribution with identical first moments. If A_n is approximated by a lognormal distribution, then one obtains an explicit formula for the approximated price resembling the Black-Scholes price of a European option. Higher moments of A_n can then be used to improve the approximation in terms of an Edgeworth series expansion (this approach is known as the Turnbull-Wakeman approximation). Another natural and usual effective choice is to approximate A_n by a distribution of the same class as X. All these approximations have been worked out in detail for the normal inverse Gaussian Lévy model in [2] and for the variance gamma Lévy model in [3]. They turn out to be a quick and accurate alternative to other numerical pricing techniques, the approximation error typically being less than 0.5% (for an extensive numerical study we refer to Albrecher and Predota [2, 3]).
Note that whereas the effectiveness of most of the other numerical techniques depends quite strongly on the structure of the marginal distributions of the Lévy process, the above approach is applicable for arbitrary risk-neutral measures and arbitrary exponential Lévy models as long as the risk-neutral moment-generating function of the log returns exists in the interval $[0, k]$.

The sensitivity of the price of an Asian option on the underlying market model has been investigated in [2, 3]). As an illustrative example, Figure 3 (taken from [2]) depicts the difference of Asian call option prices for a Black-Scholes model (in which $\mathbb{Q}$ is unique) and the Esscher price in a normal inverse Gaussian Lévy model across different strikes and maturities, where the two models were fitted to historical data of OMV daily log returns ($S_0 = 100$, daily averaging and the prices were determined by Monte Carlo simulation).

The behavior of the price difference in Figure 3 is quite typical. At the money, where most of the volume is traded, the Black-Scholes price is too high. In and out of the money, it is too low. This is intuitively clear since the Black-Scholes model underestimates the risk of larger asset price moves. If the option is very deep in or out of the money, the option price is more or less model

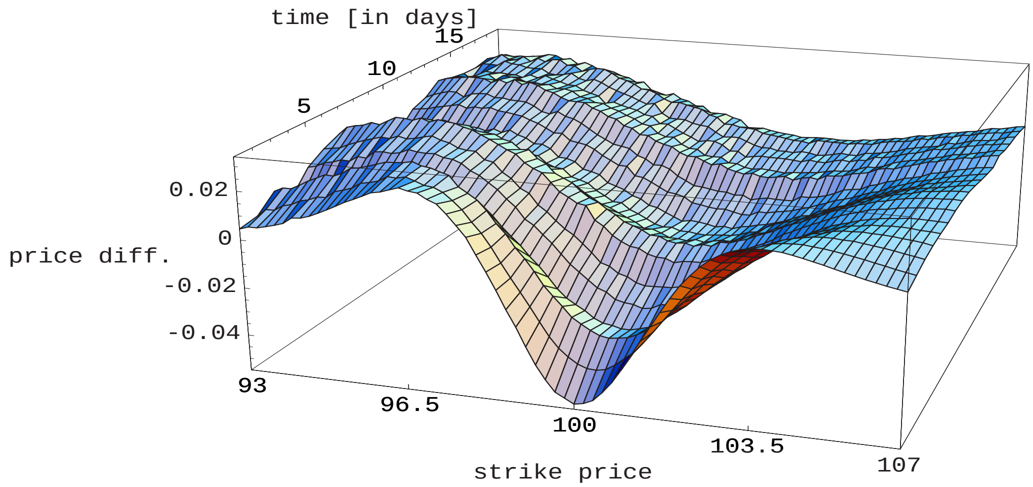


Figure 3: ${}^{(\mathrm{NIG})}\mathrm{AA}_0 - {}^{(\mathrm{BS})}\mathrm{AA}_0$ (Asian arithmetic option)

independent. The difference in option prices becomes less pronounced for increasing maturity. A comparison with the corresponding sensitivity of European option prices on the underlying model shows that the effects are quite similar, see e.g. [3, 21].

5. HEDGING OF ASIAN OPTIONS

In many circumstances the availability of a hedging strategy for a financial product is far more important than the determination of its price (note that in view of the incompleteness of the market, there exists a whole interval of no-arbitrage prices for the product depending on the particular choice of the risk-neutral measure $\mathbb{Q}$, which limits the explanatory power of a ”price”). Moreover, hedging strategies are utilized as devices for representing risk in standard reports. Even in the Black-Scholes world, hedging an Asian option is far from trivial. One approach is to derive upper and lower analytic bounds for the option price based on conditioning of random variables (for instance conditioning on the geometric average) and then to apply delta-hedging in terms of these bounds (see e.g. [30, 38]). Since these conditioning techniques are based on the simple structure of the log-normal distribution of the Black-Scholes model, it does not seem feasible to extend this approach to arbitrary exponential Lévy models. Another possibility is to apply a log-normal approximation to the dependent sum in (2) using the moment-matching technique discussed in Section 4 and then use the resulting closed-form expression of the price for delta-hedging. However, it is difficult to keep track of the implied hedging error in this case and the latter can be quite substantial since the log-normal fit of A_n may be quite poor. Moreover, delta-hedging itself is to be considered with care, since, while producing stable payoffs under idealized conditions (no limit on frequency of rehedging, no transaction costs), it produces highly variable payoffs under realistic conditions (limitations on the hedging liquidity, transaction costs). Therefore it is desirable to develop static hedging strategies where the initial hedge is kept in place for the whole lifetime of the product (or quasi-static strategies with only a small number of hedge adjustments).

In the sequel we will discuss a simple static superhedging strategy for fixed-strike Asian call options which was developed in Albrecher et al. [1]. It is based on a buy-and-hold strategy consisting of European call options maturing with and before the Asian option. For that purpose let us consider the following upper bound for the price given in (2): $\forall\ K_1, \ldots, K_n \geq 0$ with $K = \sum_{k=1}^n K_k$, we have

$$\left(\sum_{k=1}^n S_{t_k} - nK\right)^+ = \Big((S_{t_1} - nK_1) + \cdots + (S_{t_n} - nK_n)\Big)^+ \leq \sum_{k=1}^n (S_{t_k} - nK_k)^+$$

implying

$$AA_0 \leq \frac{\exp(-rT)}{n} \sum_{k=1}^n E^{\mathbb{Q}} \left[(S_{t_k} - nK_k)^+ \Big| \mathcal{F}_0 \right] = \frac{1}{n} \sum_{k=1}^n \exp(r(T-t_k)) EC_0(\kappa_k, t_k), \quad (5)$$

where $EC_0(nK_k, t_k)$ is the price of a European call option at time 0 with strike nK_k and maturity t_k. One observes that buying $\exp(-r(T-t_k))/n$ European call options at time $t = 0$ (with strike κ_k, maturity t_k) ($k = 1, \ldots, n$), holding them until their expiry and putting their payoff on the bank account represents a static superhedging strategy for this Asian option.

One still has the freedom to choose values K_k such that $\sum_{k=1}^n K_k = K$. A trivial choice is $K_k = K/n$ $(k = 1, \ldots, n)$. Since $\forall K \geq 0$ one has $EC_0(K, t) \leq EC_0(K, T)$ $(0 \leq t \leq T)$ (note that this inequality even holds if we allow for a dividend rate q as long as $q \leq r$), leading to $AA_0 \leq EC_0$, so that an Asian call option with strike K and maturity T is always dominated by a European call option with same strike and maturity. This result holds for arbitrary arbitrage-free market models; for the Black-Scholes setting it was already derived by Kemna and Vorst [27], see also [30].

Since the aim is to optimize the performance of the superhedge, one needs to determine the combination of K_k that minimizes (5). In the Black-Scholes model, this has been achieved by Nielsen and Sandmann [30] using Lagrange functions. In the general case, it turns out that comonotonicity theory leads to the optimal choice of the strike prices. Let $F(x_k; t_k) = P^{\mathbb{Q}}\left(S_{t_k} \leq x_k \,|\, \mathcal{F}_0\right)$ $(x_k, t_k > 0)$ denote the marginal distribution function of S_{t_k}. Then the optimal choice of strike prices is given by

$$n\, K_k = F^{-1}\left(F_{S^c}(n\, K); t_k\right), \quad k = 1, \ldots, n,$$

where F_{S^c} is the distribution function of the comonotone sum of $S_{t_1}, \ldots, S_{t_n}$ determined by $F_{S^c}^{-1}(x) = \sum_{k=1}^n F^{-1}(x; t_k)$. These values can be determined within less than a minute on a normal PC for the entire hedge portfolio. Note that in (2) we have $t_k = k$ $(k = 1, \ldots, n)$. Whereas the upper bound $AA_0 \leq EC_0$ (leading itself to a trivial super-hedge) is model-independent, the performance of the superhedge (5) can thus be optimized by specifying a market model and a risk-neutral measure $\mathbb{Q}$. For a proof of the optimality we refer to [1], where one can also find a numerical study of the performance of this superhedging strategy for normal inverse Gaussian, variance gamma and Meixner Lévy models (with the mean-correcting measure used for $\mathbb{Q}$). The numerical results indicate that this strategy is quite effective, in particular for low values of the strike price K. For an option with moneyness of 80%, the difference between the hedging cost and the estimated option price is typically around 1.5%, whereas the classical hedge with the European call leads to a difference of almost 10%. For options out of the money, the difference increases, but in view of the easy and cheap way in which this hedge can be implemented in practice, the comonotonic approach seems to be competitive also in these cases. Furthermore, the European call options needed for the hedge are typically available on the market and quite liquidly traded. In addition, static hedging is not exposed to the risk inherent in dynamic hedging, namely that at times of large market moves liquidity may dry up making rebalancing impossible. But especially in these situations effective hedging is needed (for further discussions on the topic, we refer to [4, 12, 15, 16]). Finally, the proposed hedging strategy works whenever an approximation of the risk-neutral density is available and can thus also be applied to stochastic volatility models using Fast Fourier transforms.

Remark: The results presented in this survey were formulated for fixed-strike arithmetic average call options. However, many of them translate immediately to put options and floating-strike options (using put-call parity and symmetries of floating and fixed strike Asian options recently established for exponential Lévy models in [23]). The inclusion of dividend payments in the model is also merely a matter of notation. Furthermore, the approximation technique of Section 4 can be adapted to geometric average rate options (cf. [2, 3]).

References

[1] Albrecher H., Dhaene J., Goovaerts M. and Schoutens W. (2003): Static hedging of Asian options under Lévy models: the comonotonicity approach. *UCS Report 2003-04*, Katholieke Universiteit Leuven.

[2] Albrecher H. and Predota M. (2004): On Asian option pricing for NIG Lévy processes. *Journal of Computational and Applied Mathematics*, to appear.

[3] Albrecher H. and Predota M. (2002): Bounds and approximations for discrete Asian options in a variance-gamma model. *Grazer Math. Ber.*, **345**, 35-57.

[4] Allen S. and Padovani O. (2002): Risk management using quasi-static hedging. *Economic Notes*, **31**(2), 277-336.

[5] Barndorff-Nielsen O.E. (1977): Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society London A*, **353**, 401-419.

[6] Barndorff-Nielsen O.E. (1998): Processes of normal inverse Gaussian type. *Finance and Stochastics*, **2**, 41-68.

[7] Benhamou E. (2002): Fast Fourier Transform for Discrete Asian Options. *Journal of Computational Finance*, **6**.

[8] Bibby M. and Sørensen M. (2003): Hyperbolic Processes in Finance. *In: Handbook of Heavy Tailed Distributions in Finance*, Rachev S. (Ed.), Elsevier Science, 211-248.

[9] Broadie M. and Glasserman P. (1996): Estimating security price derivatives using simulation. *Management Science*, **42**, 269-285.

[10] Broadie M., Glasserman P. and Kou S.G. (1999): Connecting discrete and continuous path-dependent options. *Finance and Stochastics*, **3**, 55-82.

[11] Bühlmann H., Delbaen F., Embrechts P. and Shiryaev A.N. (1996): No-arbitrage, change of measure and conditional Esscher transforms, *CWI Quarterly*,**9**, 291-317.

[12] Carr P., Ellis K. and Gupta V. (1998): Static Hedging of Exotic Options. *The Journal of Finance*, **53**, 1165-1190.

[13] Carr P., Geman H., Madan D. and Yor M. (2002): The fine structure of asset returns: an empirical investigation. *Journal of Business*, **75**, 305-332.

[14] Carr P., Geman H., Madan D.H. and Yor M. (2003): Stochastic Volatility for Lévy Processes. *Mathematical Finance*, **13**, 345-382.

[15] Carr P. and Picron J. (1999): Static Hedging of Timing Risk. *Journal of Derivatives*, **6**, 57-70.

[16] Carr P. and Wu L. (2002): Static Hedging of Standard Options. *Preprint*.

[17] Carverhill A. and Clewlow L. (1990): Flexible Convolution. *Risk*, **3**, 25-29.

[18] Chan T. (1999): Pricing contingent claims on stocks driven by Lévy processes. *Annals of Applied Probability*, **9**, 504-528.

[19] Cherny, A. (2001): No-arbitrage and completeness for the linear and exponential models based on Lévy processes, *Research Report 2001-33*, MaPhySto, University of Aarhus.

[20] Cont R. and Tankov P. (2004): *Financial Modelling with Jump Processes*. Chapman & Hall.

[21] Eberlein E. (1999): Application of generalized hyperbolic Lévy motions to finance. *In: Lévy Processes: Theory and Applications*, Barndorff-Nielsen O.E., Mikosch T. and Resnick S. (Eds.), Birkhäuser, 319-337.

[22] Eberlein E. and Özkan F. (2003): Time consistency of Lévy models. *Quantitative Finance*, **3**, 40-50.

[23] Eberlein E. and Papapantoleon A. (2003): Equivalence of Floating and Fixed Strike Asian and Lookback Options. *Preprint*, University of Freiburg.

[24] Eberlein E. and Prause K. (2001): The Generalized Hyperbolic Model: Financial Derivatives and Risk Measures. *In: Mathematical Finance - Bachelier Congress 2000*, Geman H., Madan D., Pliska S. and Vorst T. (Eds.), Springer.

[25] Gerber H. and Shiu E. (1994): Option Pricing by Esscher Transforms, *Transactions of the Society of Actuaries*, **XLVI**, 99-140.

[26] Gerber H. and Shiu E. (1994): Author's Reply to the Discussions on the paper "Option Pricing by Esscher Transforms", *Transactions of the Society of Actuaries*, **XLVI**, 175-177.

[27] Kemna A. and Vorst A. (1990): A pricing method for options based on average asset values, *Journal of Banking and Finance*, **14**, 113-129.

[28] Levy E. (1992): Pricing European average rate currency options. *Journal of International Money and Finance*, **11**, 474-491.

[29] Madan D.B. and Milne F. (1991): Option pricing with V.G. martingale components, *Mathematical Finance*, **1**, 39-55.

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

[31] Prause K. (1999): The generalized hyperbolic model: estimation, financial derivatives and risk measures, *PhD-Thesis*, University of Freiburg.

[32] Raible S. (2000): Lévy processes in Finance: Theory, Numerics and Empirical Facts, *PhD-Thesis*, University of Freiburg.

[33] Rasmus S., Asmussen S. and Wiktorsson M. (2004): Pricing of some exotic options with NIG Lévy input. *Computational Science ICCS-2004, Springer*, to appear.

[34] Rydberg T. (1999): Generalized hyperbolic diffusion processes with applications in finance, *Mathematical Finance*, **9**, 183-201.

[35] Schoutens W. (2002): The Meixner Process: Theory and Applications in Finance. *EURANDOM Report 2002-004*, EURANDOM, Eindhoven.

[36] Schoutens W. (2003): *Lévy Processes in Finance: Pricing Financial Derivatives.* Wiley, Chichester.

[37] Schoutens W. and Symens S. (2003): The pricing of exotic options by Monte-Carlo Simulations in a Lévy market with stochastic volatility. *International Journal of Theoretical and Applied Finance*, **6**(8), 839-864.

[38] Thompson G. (2000): Fast narrow bounds on the value of Asian options. *Technical Report*, University of Cambridge.
`http://mahd-pc.jims.cam.ac.uk/archive/giles/thesis/asian.pdf`

[39] Turnbull S. and Wakeman L. (1991): A Quick Algorithm for Pricing European Average Options. *Journal of Financial and Quantitative Analysis*, **26**, 377-389.

[40] Večeř J. and Xu M. (2004): Pricing Asian options in a semimartingale model. *Quantitative Finance*, **4**, 170-175.

[41] Vorst T.C.F. (1992): Prices and Hedge Ratios of Average Exchange Rate Options. *International Review of Financial Analysis*, **1**, 179-193.

HOW TO DEAL WITH CORRELATED RISKS IN ACTUARIAL SCIENCE?

A case study with Loss-ALAE data

Michel Denuit[⋆,†] **and Oana Purcaru**[⋆]

Institut de Statistique[⋆] *& Institut des Sciences Actuarielles*[†]*, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*
Email: denuit@stat.ucl.ac.be, purcaru@stat.ucl.ac.be

Abstract

This paper considers the bivariate Loss-ALAE modelling problem in actuarial science, taking into account the particular form of the censorship affecting the data. Specifically, a selection procedure for the generator of the underlying archimedean copula is described. A data set provided by the US Insurance Services Office is used for the numerical illustrations, and a comparison with previous results appeared in the actuarial literature is performed.

1. INTRODUCTION AND MOTIVATION

1.1. Losses and their associated ALAE's

Various processes in casualty insurance involve correlated pairs of variables. A prominent example is the loss and allocated loss adjustment expenses (ALAE, in short) on a single claim. Here ALAE are type of insurance company expenses that are specifically attributable to the settlement of individual claims such as lawyers' fees and claims investigation expenses.

Expensive claims generally need some time to be settled and induce considerable costs for the insurance company. Actuaries therefore expect some positive dependence between losses and their associated ALAE, i.e. large values for losses tend to be associated with large values for ALAE.

As it will be precisely explained below, the possible dependence between losses and ALAE has to be accounted for when reinsurance treaties are priced. The reinsurer covers the largest losses (i.e. those exceeding some high threshold called the retention of the direct insurer, and pays that part exceeding this threshold). It also contributes to pay the associated settlement costs on a pro rata basis. In many cases, neglecting the dependence exhibited by the data leads to serious underestimation of the expected reinsurer's payment. It is therefore crucial for the reinsurer to have at its disposal an appropriate model for the random couple Loss-ALAE.

1.2. Copula modelling for Loss-ALAE

The copula construction turns out to be very useful for the analysis of dependence between continuous outcomes. The idea behind the copula construction can be summarized as follows: for multivariate distributions, the univariate marginals and the dependence structure can be separated and the latter may be represented by a copula.

The joint modelling in parametric settings of losses and associated ALAE has been examined by FREES and VALDEZ (1998) (Pareto marginals and Gumbel copula) and KLUGMAN and PARSA (1999) (inverse paralogistic for Loss, inverse Burr for ALAE and Frank copula). Besides choosing appropriate models for the marginals, the selection of the underlying copula requires careful examination (since the dependence structure drastically affects the amount of reinsurance premiums).

1.3. Presentation of the ISO data set

The data used in the present paper were collected by the US Insurance Services Office, and comprise general liability claims randomly chosen from late settlement lags. The data consist in $n = 1,500$ observations, each accompanied by a policy limit ℓ (that is, the maximal claim amount) specific to each contract. Therefore the loss variable will be censored when the amount of claim exceeds the policy limit. More precisely, one observes a couple $(T_i, ALAE_i)$, where $T_i = \min(loss_i, \ell_i)$ for $i = 1, \ldots, n$ and an indicator

$$\delta_i = \mathbb{I}[T_i = \ell_i] = \begin{cases} 1, & \text{if} \quad loss_i > \ell_i \quad (\text{ censored claim}) \\ 0, & \text{if} \quad loss_i \leq \ell_i \quad (\text{ uncensored claim}). \end{cases}$$

Some summary statistics are gathered in Table 1. It appears clearly that, even if the vast majority of losses are uncensored (1,466 among the 1,500 observations), the 34 censored data points have a much higher mean than the 1,466 complete data (217,941 US$ versus 37,110 US$). A scatterplot of (loss, ALAE) on the log scale is depicted in Figure 1. Its shape suggests some positive relationship between loss and ALAE: large losses tend to be associated with large ALAE's, as expected. Moreover, the censored losses cluster to the right.

1.4. The need for a joint modelling of Loss-ALAE to compute reinsurance premiums

Let us now discuss practical implications of the modelling of dependence in the Loss-ALAE data. We look at the impact on premium valuation in reinsurance treaties. Let us consider a typical reinsurance treaty with limit L and insurer's retention R. Assuming a pro rata sharing of expenses, the reinsurer's payment for a given realization of (loss,ALAE) is described by the function:

$$g(\text{loss,ALAE}) = \begin{cases} 0, & \text{if } loss < R, \\ loss - R + \dfrac{loss - R}{loss}\text{ALAE}, & \text{if } R \leq loss < L, \\ L - R + \dfrac{L - R}{L}\text{ALAE}, & \text{if } loss \geq L. \end{cases}$$

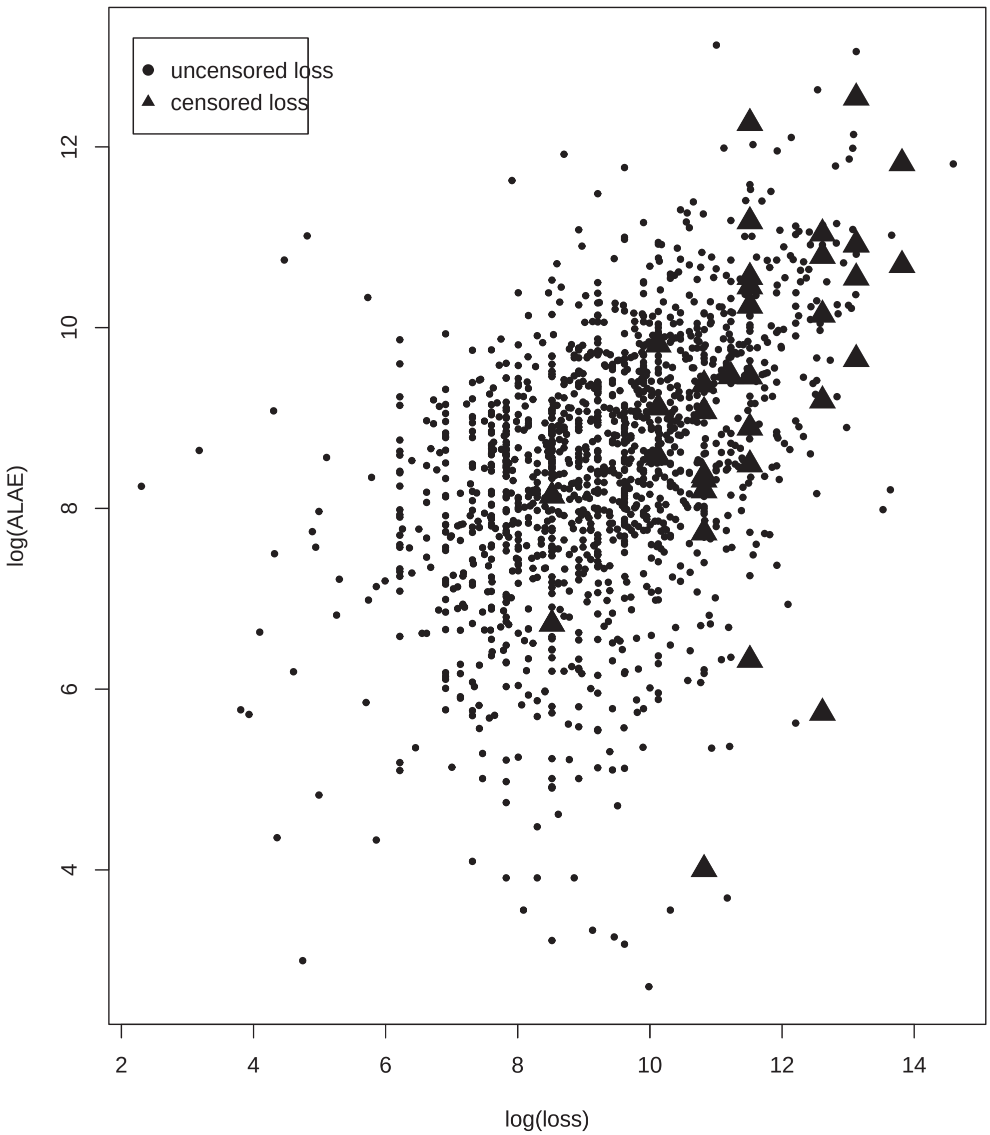


Figure 1: Scatterplot for Loss and ALAE (log-scale)).

	Loss	ALAE	Loss (uncensored)	Loss (censored)
Total N	1,500	1,500	1,466	34
Min	10 US$	15 US$	10 US$	5,000 US$
1st Qu.	4,000 US$	2,333 US$	3,750 US$	50,000 US$
Mean	41,208 US$	12,588 US$	37,110 US$	217,941 US$
Median	12,000 US$	5,471 US$	11,049 US$	100,000 US$
3rd Qu.	35,000 US$	12,577 US$	32,000 US$	300,000 US$
Max	2,173,595 US$	501,863 US$	2,173,595 US$	1,000,000
Std Dev.	102,748	28,146	92,513	258,205

Table 1: Summary statistics for variables Loss and ALAE.

To be more specific, we have Figure 2. The latter explains how a given amount of loss is divided between the policyholder i, the insurer and the reinsurer when $R \leq L \leq \ell_i$. The insurance company pays from ground up to the amount R. Then, the reinsurer covers the claim from R to L. The direct insurer then has to indemnify the policyholder from R to ℓ_i. Finally, the excess over the policy limit ℓ_i has to be supported by the policyholder. It is worth mentioning that the limit L is the same for the whole portfolio whereas the ℓ_i's are specific to each policy.

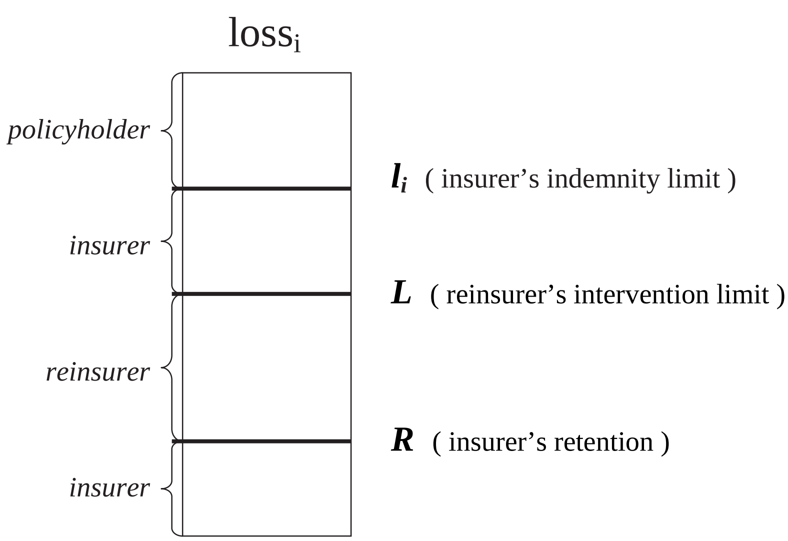


Figure 2: Splitting of the loss between reinsurer, direct insurer and policyholder.

1.5. Modelling Loss-ALAE data with archimedean copulas

A lot of recent research has focused on a subclass of copulas called the archimedean copula class, which indexes the copula by a univariate function (called the generator) and therefore yields more tractable analytical properties. Many well-known systems of bivariate distributions belong to the archimedean class. Frailty models also fall under that general prescription.

Because copulas characterize the dependence structure of a random vector once the effect of the marginals has been factored out, identifying and fitting a copula to data is not an easy task. In practice, it is often preferable to restrict the search of an appropriate copula to some rich family, like the archimedean one. Then, it is extremely useful to have simple graphical procedures to select the best fitting model among some competing alternatives for the data at hand. Starting from the assumption that the archimedean dependence structure is appropriate (an assumption that we will retain throughout this work), GENEST & RIVEST (1993) proposed such a procedure for selecting a parametric generator. Their method relies on the estimation of the univariate distribution function associated with the probability integral transformation and requires complete data. Specifically, the best fitting archimedean model is the one whose probability integral transformation distribution is closest to its empirical estimates. WANG & WELLS (2001) extended GENEST & RIVEST (1993) to right-censored bivariate failure-time data. This kind of censorship is not the one encountered in actuarial problems but, as pointed out by WANG & WELLS (2001), because the censoring issue is handled in the stage of estimating the bivariate distribution function, the approach they propose is flexible enough to deal with other censoring mechanisms. This is precisely the route we follow in this paper to deal with the modelling of Loss-ALAE.

1.6. Aim and scope of the paper

In FREES & VALDEZ (1998), techniques developed by GENEST & RIVEST (1993) for complete data have been applied to Loss-ALAE data in order to select the appropriate generator. As pointed out by these authors in their Section 4.2.1, censoring in the loss variable is ignored in the identification process. Because of censorship in the loss variable, we will develop an appropriate nonparametric estimator of the joint distribution of Loss-ALAE taking into account the particular censorship present in the data. Specifically, we follow the general approach described in WANG & WELLS (2001). Since only loss is subject to censoring, we follow the method proposed in AKRITAS (1994).

1.7. Agenda

Section 2 proposes a short tutorial about copulas. In Section 3, we propose a new nonparametric estimator for the generating function, that takes into account the fact that losses may be censored whereas ALAE's are complete. This nonparametric estimation then serves as a benchmark to select an appropriate parametric archimedean copula. The paper ends with numerical illustrations, given in Section 4.

2. ARCHIMEDEAN COPULAS

2.1. Sklar's theorem

Broadly speaking, a copula is (the restriction to the unit square $[0,1]^2$ of) a joint cdf for a bivariate random vector with unit uniform marginals. More formally, a copula C is a function mapping the unit square $[0,1]^2$ to the unit interval $[0,1]$ which is non-decreasing and right-continuous, and satisfies

(i) $C(0,u_2) = C(u_1,0) = 0$;

(ii) $C(u_1,1) = u_1$ and $C(1,u_2) = u_2$;

(iii) $C(v_1,v_2) - C(u_1,v_2) - C(v_1,u_2) + C(u_1,u_2) \geq 0$ for any $u_1 \leq v_1$, $u_2 \leq v_2$.

Sklar's theorem elucidates the role that copulas play in the relationship between multivariate cdf's and their univariate margins. Specifically, given a bivariate cdf $F_{\boldsymbol{X}}$ with univariate marginal cdf's F_1 and F_2, there exists a copula C such that for all $\boldsymbol{x} \in \mathbb{R}^2$ the joint cdf $F_{\boldsymbol{X}}$ can be represented as:

$$F_{\boldsymbol{X}}(x_1,x_2) = C\left(F_1(x_1), F_2(x_2)\right), \quad (x_1,x_2) \in \mathbb{R}^2. \tag{1}$$

When the marginals F_1 and F_2 are continuous, then the copula C in (1) is unique. Otherwise C is uniquely determined on Range$(F_1)\times$Range(F_2). Conversely, if C is a copula and F_1 and F_2 are cdf's then the function $F_{\boldsymbol{X}}$ defined by (1) is a bivariate cdf with margins F_1 and F_2. The explicit expression for the copula C when the marginals are continuous is

$$C(\boldsymbol{u}) = F_{\boldsymbol{X}}\big(F_1^{-1}(u_1), F_2^{-1}(u_2)\big), \;\; (u_1,u_2) \in [0,1]^2. \tag{2}$$

Formal proofs can be found e.g. in NELSEN (1999).

2.2. Archimedean family

Consider a twice-differentiable strictly decreasing and convex function $\phi : [0,1] \to [0,+\infty]$ satisfying $\phi(1) = 0$. These requirements are enough to guarantee that ϕ has an inverse ϕ^{-1} having also two derivatives. Every such function ϕ generates a bivariate distribution function C_ϕ whose marginals are uniform on the unit interval (i.e. a copula) given by

$$C_\phi(u_1,u_2) = \begin{cases} \phi^{-1}\left\{\phi(u_1)+\phi(u_2)\right\} & \text{if } \phi(u_1)+\phi(u_2) \leq \phi(0), \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

for $0 \leq u_1, u_2 \leq 1$. Copulas C_ϕ of the form (3) are referred to as Archimedean copulas. The function ϕ is called the generator of the copula. Only ϕ functions satisfying $\lim_{t\to 0+}\phi(t) = +\infty$ are used in this work. This ensures that C_ϕ is absolutely continuous.

Now, a bivariate distribution function $F_{\boldsymbol{X}}$ with marginals F_1 and F_2 is said to be generated by an Archimedean copula if, and only if (1) holds with an Archimedean copula C_ϕ.

2.3. Some members of the archimedean family

2.3.1. Clayton copula

For $\alpha > 0$, Clayton copula is given by

$$C_\alpha(u_1, u_2) = \left(u_1^{-\alpha} + u_2^{-\alpha} - 1\right)^{-1/\alpha}.$$

It is the archimedean copula associated to the generator

$$\phi_\alpha(t) = t^{-\alpha} - 1, \ \ \alpha > 0.$$

The parameter α can be interpreted as a measure of the strength of the dependence. It can be shown that the association between the components increases with α in the concordance order.

2.3.2. Frank copula

Frank's copula is given by

$$C_\alpha(u_1, u_2) = -\frac{1}{\alpha} \ln\left(1 + \frac{(\exp(-\alpha u_1) - 1)(\exp(-\alpha u_2) - 1)}{\exp(-\alpha) - 1}\right), \ \ \alpha \neq 0.$$

It is the archimedean copula associated to

$$\phi_\alpha(t) = -\ln \frac{\exp(-\alpha t) - 1}{\exp(-\alpha) - 1}, \ \ \alpha \in \mathbb{R}.$$

The larger α in absolute value, the stronger the association (as measured by the concordance order). A positive (resp. negative) value of α indicates positive (resp. negative) dependence.

2.3.3. Gumbel copula

Gumbel copula has the form

$$C_\alpha(u_1, u_2) = \exp\left(-\left\{(-\ln u_1)^\alpha + (-\ln u_2)^\alpha\right\}^{1/\alpha}\right), \ \ \alpha \geq 1,$$

It is the archimedean copula associated with

$$\phi_\alpha(t) = \Big(-\ln(t)\Big)^\alpha, \ \ \alpha \geq 1.$$

The parameter α controls the amount of dependence between the two components (in the concordance order).

This copula is consistent with bivariate extreme value theory and can be used to model the limiting dependence structure of componentwise maxima of bivariate random couples.

2.4. Bivariate probability integral transformation theorem

It is well-known that given any random variable X with continuous distribution function F, $F(X)$ is uniformly distributed on the interval $[0,1]$. This fundamental result is known as the Probability Integral Tranformation (PIT) theorem and underlies many statistical procedures. In particular, when F_1 and F_2 are continuous, the copula C for (X_1,X_2) is just the joint cdf for $\big(F_1(X_1),F_2(X_2)\big)$.

Now, if we define the bivariate PIT of (X_1,X_2) as $Z=F_{\boldsymbol{X}}(X_1,X_2)$, it is not generally true that the cdf K of Z is uniform on $[0,1]$, even when $F_{\boldsymbol{X}}$ is continuous. Moreover, K does not characterize $F_{\boldsymbol{X}}$ since K does not contain any information about the marginals F_1 and F_2 because

$$Z=F_{\boldsymbol{X}}(X_1,X_2)=C(U_1,U_2)$$

where (U_1,U_2) admits C as joint cdf.

Let us now examine the bivariate PIT for archimedean copulas. The following result is due to GENEST & RIVEST (1993).

Proposition 2.1 *Let U be a random couple with unit uniform marginals and joint cdf C_ϕ. Let us define $Z=C_\phi(U_1,U_2)$. The cdf K of Z is given by $K(z)=z-\lambda(z)$ where*

$$\lambda(\xi)=\frac{\phi(\xi)}{\phi^{(1)}(\xi)},\ \ 0<\xi\leq 1.$$

3. SELECTING THE GENERATOR WITH CENSORED DATA

3.1. Nonparametric estimation of the generator

The nonparametric estimation procedure of the generator is based on the fact that it is possible to estimate K nonparametrically from a random sample of (X_1,X_2) pairs. This provides for archimedean copulas an indirect way of estimating the generator ϕ (and hence the copula C_ϕ) by virtue of Proposition 2.1. Indeed, given K, it is possible to recover ϕ by solving the differential equation

$$\frac{\phi(z)}{\phi^{(1)}(z)}=z-K(z)$$

which yields

$$\phi(z)=\exp\left\{\int_{\xi=z_0}^{z}\frac{1}{\xi-K(\xi)}d\xi\right\} \tag{4}$$

where $0<z_0<1$ is an arbitrary chosen constant (coming back to (3), it is easily seen that ϕ is defined up to a positive factor). The function ϕ defined in (4) generates an archimedean copula whenever $z-K(z)$ is negative and remains bounded away from 0 on the unit interval. Specifically, GENEST & RIVEST (1993) derived the following result.

Proposition 3.1 *The function ϕ given in* (4) *is decreasing and convex and satisfies $\phi(1)=0$ if, and only if,*

$$K(z-)=\lim_{t\to z-}K(z)>z\quad \textit{for all}\quad 0<z<1.$$

The condition involved in Proposition 3.1 has to be fulfilled by the estimator of K in order to recover a bona fide generator from (4). More specifically, under the assumption that the dependence function associated with K is archimedean, a natural estimator $\widehat{\lambda}_n$ of λ can be derived from an estimator $\widehat{K}_n$ of K through the relation

$$\widehat{\lambda}_n(z) = z - \widehat{K}_n(z), \;\; 0 < z < 1.$$

Provided $\widehat{K}_n(z-) > z$ for all $0 < z < 1$, formula (4) then provides an estimator of C_ϕ within the class of archimedean copulas.

3.2. Genest-Rivest estimation procedure for the generator in the presence of complete data

GENEST and RIVEST (1993) were the first to propose a procedure for identifying a generator in empirical applications. Given observations from a random pair $\boldsymbol{X} = (X_1, X_2)$ with cdf $F_{\boldsymbol{X}}$, this procedure relies on the estimation of the univariate cdf associated with the probability integral transformation $Z = F_{\boldsymbol{X}}(X_1, X_2)$. If the data were complete, $K(\cdot)$ could be estimated by the empirical cdf of the pseudo-observations

$$z_i = \frac{1}{n-1}\#\left\{(x_1^{(j)}, x_2^{(j)})\middle| x_1^{(j)} < x_1^{(i)}, x_2^{(j)} < x_2^{(i)}\right\}$$

i.e.

$$\widehat{K}_n(z) = \frac{1}{n}\#\{i|z_i \leq z\}$$

where the symbol $\#$ stands for the cardinality of a set.

3.3. Wang-Wells estimation procedure for the generator in the presence of censored data

However this technique is no longer appropriate when the data is subject to censoring. For such cases, WANG & WELLS (2000) propose a modified estimator of $K(\cdot)$. Since $K(\cdot)$ can be written as:

$$K(z) \;\; = \;\; \Pr[F_{\boldsymbol{X}}(X_1, X_2) \leq z] = \mathbb{E}[\mathbb{I}\{F_{\boldsymbol{X}}(X_1, X_2) \leq z\}]$$

the suggested estimator is given by :

$$\hat{K}_n(z) = \int_0^\infty \int_0^\infty \mathbb{I}[\hat{F}_{\boldsymbol{X}}(x_1, x_2) \leq z] d\hat{F}_{\boldsymbol{X}}(x_1, x_2) \tag{5}$$

where $\hat{F}_{\boldsymbol{X}}$ stands for a nonparametric estimator of the joint distribution function $F_{\boldsymbol{X}}$ taking censoring into account. As mentioned by the authors, this approach is sufficiently flexible to deal with various censorship mechanisms, as long as $\hat{F}_{\boldsymbol{X}}$ is an appropriate estimator for $F_{\boldsymbol{X}}$.

3.4. Akritas estimation procedure for a bivariate cdf under random censoring

A nonparametric estimator for the joint distribution function, when only one variable is subject to censoring, was given by AKRITAS (1994). We will adapt this method for the case where the censoring variable is a constant specific to each observation (making the data non identically distributed).

Let (X, Y) be a couple of random variables with joint distribution function F, where X is subject to censoring, and L is the censoring variable. One observes $(T, Y) = (\min(X, L), Y)$ and an indicator $\delta = \mathbb{I}(T = L) = \mathbb{I}(X > L)$. Assume that X and L are independent (more generally, Y is supposed to be independent of L given X, but L is allowed to depend on X). The proposed estimator of F is based on the estimator of the conditional distribution $F_{1|2}(x|y) = \Pr[X \leq x | Y = y]$:

$$F(x, y) = \Pr[X \leq x, Y \leq y] = \int_0^y F_{1|2}(x|z) dF_Y(z)$$

AKRITAS (1994) showed the consistency and efficiency of an estimator based on the previous relation. Let H be a known probability density function (kernel) and $\{h_n\}$ a sequence of positive constants tending to zero as n tends to infinity (bandwidth sequence). The proposed estimator is given by :

$$\hat{F}(x, y) = \int_0^y \hat{F}_{1|2}(x|z) d\hat{F}_Y(z)$$

where $\hat{F}_Y(\cdot)$ is the empirical distribution of Y, given by

$$\hat{F}_{1|2}(x|z) = 1 - \prod_{T_i \leq x;\ \delta_i = 0} \left[1 - \frac{W_{n2i}(z; h_n)}{\sum_{j=1}^n W_{n2j}(z; h_n)\ \mathbb{I}[T_j \geq T_i]} \right]$$

with

$$W_{n2i}(y; h_n) = \frac{H\left(\frac{y - Y_i}{h_n}\right)}{\sum_{j=1}^n H\left(\frac{y - Y_j}{h_n}\right)}.$$

The estimator of the joint cdf is given by

$$\hat{F}(x, y) = \int_0^y \hat{F}_{1|2}(x|z) d\hat{F}_Y(z) = \frac{1}{n} \sum_{k=1}^n \mathbb{I}[0 \leq Y_k \leq y] \cdot \hat{F}_{1|2}(x|Y_k).$$

Now, coming back to (5), the latter estimator of the joint cdf F gives an estimator $\widehat{K}_n$ of K, yielding in turn an estimated generator ϕ via (4) provided the condition of Proposition 3.1 is fulfilled.

4. APPLICATION TO Loss-ALAE

4.1. Estimation of the generator

The bandwidth h involved in Akritas estimation procedure is selected so to minimize AMSE. The estimation of K then follows from (5); Figure 3 depicts the resulting $\widehat{K}_n$.

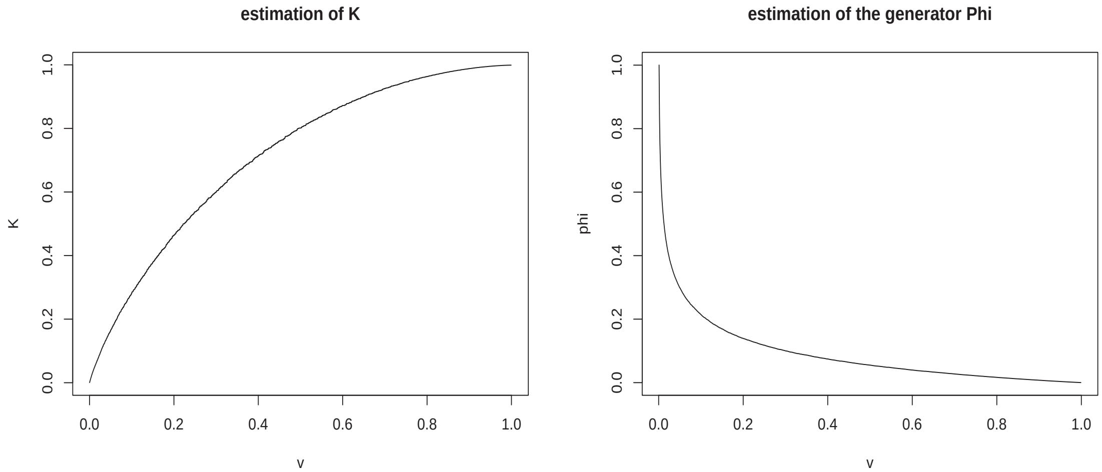


Figure 3: Graph of $\widehat{K}_n$ and resulting $\widehat{\phi}$ for the Loss-ALAE data set.

The condition of Proposition 3.1 is fulfilled. The generator of the archimedean copula is then obtained from (4), that is

$$\hat{\phi}(u) = \exp\left\{\int_{z_0}^{z} \frac{1}{t - \hat{K}(t)} dt\right\}, \qquad \text{with } u_0 = 1/1000$$

The resulting $\widehat{\phi}$ is depicted in Figure 3.

4.2. Graphical model selection procedure for Loss-ALAE

The idea is now to compare $\widehat{K}_n$ with several parametric analogues K_α corresponding for instance to Clayton, Gumbel or Frank copulas. To this end an estimation of α is needed. A convenient way to estimate the dependence parameter α is to relate it to Kendall's tau and to deduce $\widehat{\alpha}$ from $\widehat{\tau}$. Kendall's tau is easily estimated since provided the marginal distributions are continuous

$$\tau(X_1, X_2) = 4\mathbb{E}[Z] - 1 = 4\int_0^1 \big(1 - K(z)\big)dz - 1. \tag{6}$$

Since the estimation of K takes into account the censoring mechanism, the estimated τ obtained from (6) is suitable for censored data.

Of course, other approaches are possible. We use here the maximum pseudo-likelihood procedure known as omnibus. The omnibus semiparametric procedure treats marginal distributions as (infinite dimensional) nuisance parameters. This simple procedure consists in substituting empirical analogues for the marginal distribution functions in the likelihood for the dependence parameters and then in maximizing the resulting peudo-likelihood. As shown by GENEST ET AL. (1995), the resulting estimator is consistent and asymptotically normal, even in the presence of censorship.

The 2-stage semiparametric estimation assumes that the marginal distributions are unknown. First they will be estimated nonparametrically, by the Kaplan-Meier estimator (for the Loss variable) and the empirical estimator (for the ALAE variable). These estimators, $\hat{F}_X$ and $\hat{F}_Y$ respectively, are then used to estimate the dependence parameter

$$\hat{\alpha} = \arg\max_{\alpha} L(\alpha, \hat{F}_X, \hat{F}_Y)$$

Since the loss variable is censored, the likelihood function can be written as :

$$L(\alpha, u, v) = \prod_{i=1}^{n} c_\alpha(u_i, v_i)^{\delta_i} \left[1 - \frac{\partial C_\alpha(u_i, v_i)}{\partial v_i}\right]^{1-\delta_i}$$

where $(u, v) = (\hat{F}_X(x), \hat{F}_Y(y))$, C_α is the archimedean copula and c_α its density. The log-likelihood will therefore be given by :

$$\ln L(\alpha, u, v) = \sum_{i=1}^{n} \left[\delta_i \ln(c_\alpha(u_i, v_i)) + (1 - \delta_i) \ln\left(1 - \frac{\partial C_\alpha(u_i, v_i)}{\partial v_i}\right)\right]$$

The derivatives of Gumbel, Frank and Clayton's copulas, appearing in the expression of the likelihood function are given in the next table:

copula	$\frac{\partial C_\alpha(u,v)}{\partial v}$
Gumbel	$v^{-1} \cdot \exp\{-(\tilde{u}^\alpha + \tilde{v}^\alpha)^{1/\alpha}\} \cdot [1 + (\frac{\tilde{u}}{\tilde{v}})^\alpha]^{-1+1/\alpha}$
Frank	$[e^{-\alpha v} - e^{-\alpha(u+v)}] \cdot [(1 - e^{-\alpha}) - (1 - e^{-\alpha u})(1 - e^{-\alpha v})]^{-1}$
Clayton	$[1 + v^\alpha(u^{-\alpha} - 1)]^{-1-1/\alpha}$

The estimations of the dependence parameters for the three copulas mentioned before, obtained using this procedure are given in the following table:

copula	$\hat{\alpha}$
Gumbel	1.444
Frank	3.077
Clayton	0.517

In order to select the parametric form for ϕ, it suffices to compare each parametric estimate to the nonparametric estimate constructed above. The idea is to select ϕ so that the parametric estimate resembles the nonparametric one. Figure 4 displays semiparametric and parametric estimation of λ. In order to evaluate the agreement between the semiparametric estimator $\widehat{K}_n$ and parametric analogues $K_{\hat{\alpha}}$, QQ-plots are displayed in Figure 5. Measuring closeness can be done by minimizing a distance such as

$$S(\hat{\alpha}) = \int_0^1 (K_{\hat{\alpha}}(z) - \hat{K}(z))^2 dz.$$

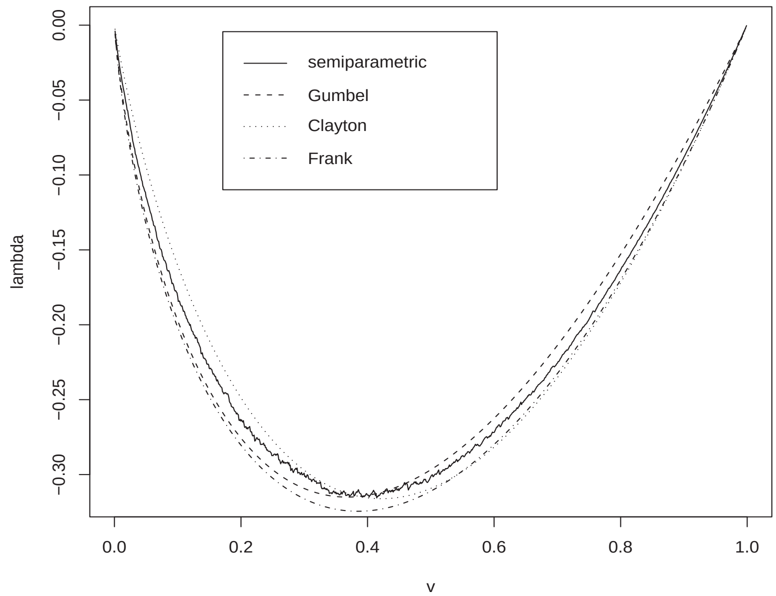


Figure 4: Semiparametric and parametric estimation of λ involved in Proposition 2.1.

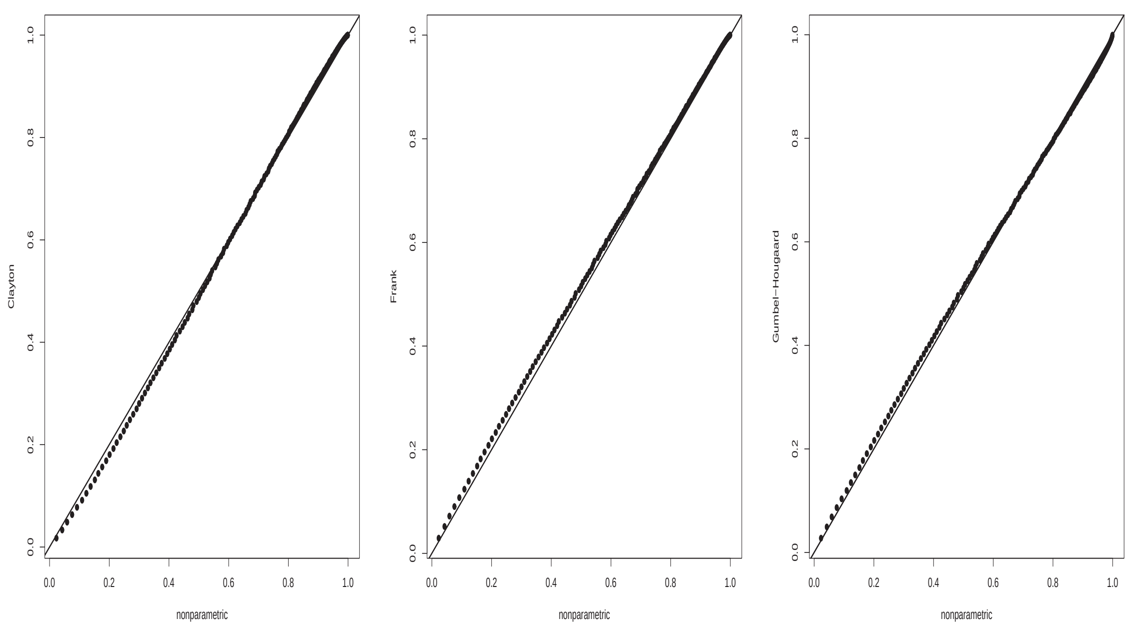


Figure 5: QQ-plots for the semiparametric $\widehat{K}_n$ and parametric analogues $K_{\widehat{\alpha}}$ associated to various archimedean models.

Here, we get

copula	$S(\hat{\alpha})$
Clayton	0.0001123993
Gumbel	9.302016e-05
Frank	0.0001477749

so that Gumbel copula is the closest to the semiparametric archimedean model.

Once a parametric model for the copula is selected, all the analyses performed by FREES ET AL. (1997) and KLUGMAN & PARSA (1999) can be replicated to the data set at hand.

Acknowledgements

We would like to thank Professors Frees and Valdez for kindly providing the Loss-ALAE data, which were collected by the US Insurance Services Office (ISO).

References

[1] Akritas M.G. (1994): Nearest neighbor estimation of a bivariate distribution under random censoring *Annals of Statististics*, **22**, 1299-1327.

[2] Frees E.W. and VALDEZ E.A. (1998): Understanding relationships using copulas. *North American Actuarial Journal*, **2**, 1-25.

[3] Genest C., Ghoudi K. and Rivest L.-P. (1995): A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, **82**, 543-552.

[4] Genest C. and Rivest L. (1993): Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, **88**, 1034-1043.

[5] Klugman S.A. and Parsa R. (1999): Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, **24**, 139-148.

[6] Nelsen R.B. (1999): *An Introduction to Copulas*. Springer-Verlag, New York.

[7] Wang W. and Wells M.T. (2000): Model selection and semiparametric inference for bivariate failure-time data. *Journal of the American Statistical Association*, **95**, 62-72.

CONTRIBUTED TALKS

COULD WE BETTER ASSESS THE RISK OF A CREDIT PORTFOLIO?

Jessica Cariboni, Dirk Baur and Francesca Campolongo

European Commission, Joint Research Center, Ispra (VA), Italy
Email: `jessica.cariboni@jrc.it`, `dirk.baur@jrc.it`,
`francesca.campolongo@jrc.it`

Abstract

Aim of this work is to show the potential use of a global sensitivity analysis to assess the riskiness of credit risk portfolios. We describe the obligors behavior via a latent factor model and we study the sensitivity of commonly used risk measures with respect to three key input factors.
Results show that a global approach provide the risk modeler with a broad picture of the risk contributions of different elements to the model. The main finding is that the obligors default probabilities and the correlation of the latent variables describe most of the volatility of the risk associated with the portfolio under investigation.

1. THE MODEL

In this work we model and analyze the risk associated to a credit portfolio. A crucial point in the model is the dependence of the obligors. The latent variables approach is chosen to describe the behavior of the m obligors (e.g. see Bluhm et al. (2003) and Frey et al. (2001)) while their dependence is modelled via the dependence of m underlying latent variables $\mathbf{W} = (W_1, \ldots, W_m)'$. Each latent variable is driven by one common factor Z and an idiosyncratic shock ϵ_j as it follows:

$$W_j = \sqrt{a_j} Z + \sqrt{1 - a_j} \epsilon_j \qquad \text{for} \qquad j = 1, 2, \ldots, m. \tag{1}$$

a_j ($\in (0,1)$) describes the exposure of obligor j to factor Z. Z and ϵ_j are assumed to be independent and identically distributed with mean zero and variance one.

Default for obligor j is described by a state indicator Y_j which takes just two values, zero and one, correspondent respectively to non default and default states. The occurrence of default for obligor j depends on a deterministic cutoff point D_j as it follows:

$$Y_j = 1 \Longleftrightarrow W_j \leq D_j.$$

The probability of default π_j is then given by:

$$\pi_j = P(W_j \leq D_j), \tag{2}$$

while the joint default probability π_{ij} for two obligors i and j can then be written as:

$$\pi_{ij} = P(W_i \leq D_i, W_j \leq D_j)$$

The loss distribution is obtained via Monte Carlo (MC) simulation as follows:

1. a MC loop is performed to determine the distribution W_j for each obligor, obtained as a function of the assigned factor loading via equation (1);

2. from the obtained W_j and the input default probabilities, the cutoff vector of the obligors $\boldsymbol{D}$ is derived by inverting equation (2);

3. from the multivariate distribution $\boldsymbol{W}$ a number of I draws is randomly selected and compared with the cutoff vector $\mathbf{D}$, to obtain the distribution of joint defaults.

2. UNCERTAINTY ANALYSIS AND SENSITIVITY ANALYSIS

A model represents a formal way to map some information and assumptions into inference. Uncertainty analysis (UA) and sensitivity analysis (SA) help the modeler in understanding the uncertainty affecting the output variable under investigation. In fact UA quantifies the volatility in the model output while SA assesses the relative importance of the input factors in determining such an uncertainty.

Suppose that the model under study maps the k-dimensional input space in the output space through a mathematical function f:

$$Y = f(X_1, \ldots, X_k).$$

Each input factor X_i can be considered as a random variable characterized by a specific probability density function (pdf). In the same way the output Y (which is here considered a scalar quantity but could also be a vector) can be thought as a random variable whose pdf is subject of investigation for UA. The output pdf is studied empirically and obtained via MC simulation.

Once the volatility of the output has been quantified, SA helps in understanding how each input affects the uncertainty in the output. Very often SA is defined as a local measure of the effect of a given input on the output (e.g. see Frey et al. (2001)), obtained by estimating system derivatives such as:

$$S_j = \frac{\partial Y}{\partial X_j}.$$

This local approach is practicable only when the variation around a fixed point of the input factors is small, or when the input-output relationship is assumed to be linear.

When the ranges of variation of the input factors are material and/or the model is non-linear the use of a global approach (Saltelli et al. (2000), (2004)) that estimates the effect of a single factor while all the others are varied as well is compulsory (see Saltelli (1999)).

In our exercise, we have no reason to believe a priori that the model is linear neither can we restrict our attention to small ranges of variation for the input factors. Therefore we propose a global sensitivity approach to determine what input factors in the portfolio model play a major role and should be modelled carefully.

2.1. The global sensitivity measures

Assume our goal is to rank the factors according to the amount of output unconditional variance $V(Y)$ that is removed when we learn the true value of a given input factor. In other words, we are facing the problem setting known in sensitivity analysis as "factor prioritization" (FP) (see Saltelli et al. (2004) and Saltelli and Tarantola (2002)).

This means that factors could be ranked according to $V(Y \mid X_i = x_i^*)$, the variance obtained by fixing X_i to its true value x_i^*. Since the true value for each input X_i is not known it sounds sensible to look at the weighted average of the above measure over all possible values x_i^* of X_i, i.e. to $E(V(Y \mid X_i))$. The smaller is $E(V(Y \mid X_i))$, the more influential is the factor X_i. Since

$$V(Y) = E(V(Y \mid X_i)) + V(E(Y \mid X_i))$$

higher values for $V_i = V(E(Y \mid X_i))$ correspond to influential factors. The quantity V_i normalized by the value of the unconditional variance is called first order sensitivity index and is used as a measure of sensitivity (Sobol' (1990) and (1993)):

$$S_i = \frac{V_i}{V}.$$

It can be demonstrated that $S_i, i = 1, 2, \ldots, k$ are the proper measure to rank the factors in order of importance in the FP setting, also in the presence of interactions[1] (see Saltelli and Tarantola (2002)).

Nevertheless when interactions are part of the model under investigation first order sensitivity coefficients are not capable to explain the entire variance of the output. This can be seen in the context of the general variance decomposition scheme proposed by Sobol' (1990, 1993) for independent input factors:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \ldots + V_{12\ldots k} \tag{3}$$

where, for instance,

$$V_{ij} = V(E(Y \mid X_i, X_j)) - V_i - V_j$$

and $S_{ij} = V_{ij}/V$ measures the interactions between X_i and X_j. Similar equations hold for higher order coefficients.

In this scheme a second sensitivity measure, the total index ST_i, can be introduced to estimate the total contribution to the variance of Y due to the factor X_i (Homma and Saltelli (1996)). The total index ST_i is defined as the sum of all the terms of the variance decomposition (3) where at

[1]Two or more factors are said to interact when their effect on Y cannot be expressed as the sum of their single effects.

least one of the indices is equal to i. It can be demonstrated that the total indices are the measures to identify unimportant factors i.e. those factors that can be fixed at any given value within their range of variation without significantly affecting the total output variance (chapter 5 in Saltelli et al. (2004)). S_i and ST_i are estimated via MC simulation.

2.2. Steps of the analysis

In the light of the above the first step of the procedure is concerned with the choice of the output variables of interest. As already pointed out we are interested in quantifying the maximal risk associated with a certain portfolio: the output variables of interest are five fixed quantiles of the loss distribution (from 95% up to 99.5%).

In the second step we choose the input factors and assign a specific pdf. The three independent factors are:

1. A trigger factor which defines the shape of the multivariate distribution of the latent factors $\mathbf{W}$. The trigger factor may assume three possible values: a Gaussian distribution, a t-distribution with 10 degrees of freedom or a t-distribution with 4 degrees of freedom.

2. A trigger factor taking on five values which determines the degree of correlation among the obligors, represented by a m-dimensional vector of a_j. The five possible vectors of loadings are randomly generated a priori under the assumption that the a_j are uniformly distributed within the following fixed ranges:

   - very low correlation i.e. $a_j \sim U[0, 0.15], j = 1, 2, \ldots, m$
   - low correlation i.e. $a_j \sim U[0.15, 0.30], j = 1, 2, \ldots, m$
   - medium-low correlation i.e. $a_j \sim U[0.30, 0.45], j = 1, 2, \ldots, m$
   - medium correlation i.e. $a_j \sim U[0.45, 0.60], j = 1, 2, \ldots, m$
   - medium-high correlation i.e. $a_j \sim U[0.60, 0.75], j = 1, 2, \ldots, m$

3. A trigger factor which determines the rating portfolio composition, represented by a m-dimensional vector of default probabilities π_j. Three possible determinations of this vector are randomly generated a priori under the assumption that the π_j are uniformly distributed within the following fixed ranges:

   - high rated obligors (e.g. class A) i.e. $\pi_j \sim U[0, 0.05], j = 1, 2, \ldots, m$
   - medium rated obligors (e.g. class B) i.e. $\pi_j \sim U[0.05, 0.10], j = 1, 2, \ldots, m$
   - low rated obligors (e.g. class C) i.e. $\pi_j \sim U[0.10, 0.15], j = 1, 2, \ldots, m$

After the definition of the input factors, an input sample of size N is generated through the use of the Sobol' method (Sobol' (1993)) and the distribution of joint defaults is computed for each sample point following the procedure described in section 1. This produces N determinations for each quantiles allowing to obtain the empirical distributions for the outputs used for UA/SA.

3. RESULTS

UA/SA results are computed for a portfolio of 1000 obligors. $I = 10.000$ is the number of draws to compute the distribution of joint defaults while $N = 16.384$ is the size of the input sample for UA/SA.

	Quantile 95%	Quantile 97.5%	Quantile 99%	Quantile 99.5%	Quantile 99.9 %
Mean	303	397	504	569	671
Std. Dev.	172	208	234	243	246
Minimum	52	60	65	73	75
Maximum	795	939	989	999	1000

Table 1: Basic statistics relative to the simulated distributions of the number of joint defaults correspondent to quantiles 95%, 97.5% , 99% , 99.5% and 99.9% for a portfolio of 1000 obligors.

The main statistics of the empirical distribution for the outputs are listed in Table 1. The numbers point out that the average number of joint defaults is rather different for the selected quantiles and that the variability of results is also pronounced.

	Quantile 95%		Quantile 97.5%		Quantile 99%		Quantile 99.5%		Quantile 99.9%	
	S	ST	S	ST	S	ST	S	ST	S	ST
Degree Correlation	0.237	0.318	0.369	0.441	0.499	0.558	0.572	0.632	0.634	0.761
Multivariate Distr.	0.021	0.035	0.026	0.050	0.033	0.078	0.049	0.107	0.065	0.178
Ptf. Composition	0.672	0.774	0.547	0.629	0.398	0.461	0.329	0.396	0.198	0.308

Table 2: Global sensitivity analysis results: first order indices S and total indices ST for the three considered input factors.

Results of global sensitivity analysis for the portfolio under investigation are presented in Table 2 where the obtained first and total order sensitivity indices are listed for the three factors in correspondence to the five outputs. The following conclusions can be easily drawn from these numbers:

1. Since the differences between S_i and ST_i are not so pronounced for each of the three factors, interactions between factors are not very important.

2. More than 80% of the total variance can be explained by the degree of correlation among obligors and by the portfolio rating composition in all cases (the sum of their first order indices is always greater than 0.8). The relative importance of the portfolio rating composition and the degree of correlation among the obligors depends strongly on the quantile that is considered. At lower quantiles the rating portfolio composition is predominant ($> 65\%$ of the total variance) while it decreases to 19.8% at higher quantiles.

3. Although influencing the number of joint defaults, the multivariate distribution of the latent variables explains a smaller fraction of the total variability in the outputs than the other two

factors. Even if we look at its overall effects (ST_i) and we consider extreme events quantile 99.9% its contribution to the total variance is lower than 20%.

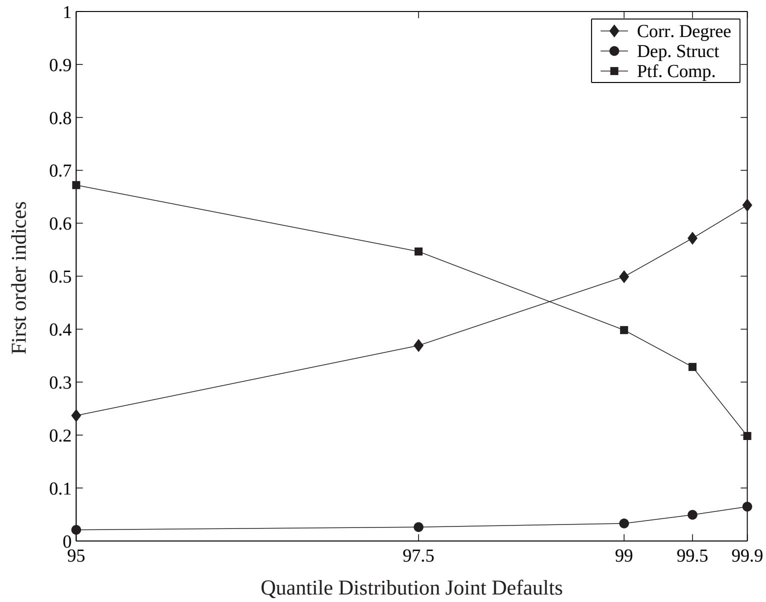


Figure 1: Evolution of first order indices for the three factors as a function of the quantiles of the loss distribution.

Figure 1 plots the evolution of first order sensitivity indices of the three factors as a function of the quantile fixed in the loss distribution. The plot shows that the shape of the multivariate distribution is much less important than the other factors at all quantiles, while the portfolio rating composition is the most influential factor for quantiles lower than $\approx 98\%$.

4. CONCLUSIONS

This paper has introduced the concept of global sensitivity analysis to evaluate credit risk models. We showed that global sensitivity analysis reveals the relative contributions of different input factors to the total output variability. Our study reveals that the portfolio composition is the most important factor for lower quantiles of the loss distribution while the degree of correlation has more influence at higher quantiles. Results show that the importance of the shape of the multivariate distribution of the latent variables is smaller than that of the other two factors.

In practice the obtained results imply that at higher quantiles, modelling the degree of correlation is more effective in reducing the uncertainty in the output than modelling the portfolio composition. More care must be placed in the credit portfolio construction at lower quantiles.

References

[1] Bluhm Ch., Overbeck L. and Wagner Ch. (2003): *An Introduction to Credit Risk Modeling*. Chapman & Hall/ CRC.

[2] Frey R., McNeil A.J. and Nyfeler M. (2001): Copulas and credit models. *RISK*, October, 111-114.

[3] Homma T. and Saltelli A. (1996): Importance measures in global sensitivity analysis of model output. *Reliability Engineering and System Safety*, **52**(1), 1-17.

[4] Saltelli, A. (1999): Sensitivity analysis. Could better methods be used? *Journal of Geophysical Research*, **104**(D3), 3789-3793.

[5] Saltelli A., Chan K. and Scott E. M. (2000): *Sensitivity Analysis*. John Wiley & Sons publishers, Probability and Statistics series.

[6] Saltelli A., Tarantola S., Campolongo F. and Ratto M. (2004): *Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models*. John Wiley & Sons publishers, Probability and Statistics series.

[7] Saltelli A. and Tarantola S. (2002): On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of the American Statistical Association*, **97**(459), 702-709.

[8] Sobol' I.M. (1990, in Russian), (1993, in English): Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modelling and Computational Experiments*, **1**(4), 407-414.

A BIMODAL VASICEK SHORT RATE MODEL

Marc Decamps[†], Marc Goovaerts[‡] and Wim Schoutens[§]

[†] *K.U.Leuven, FETEW, Naamsestraat 69, B-3000 Leuven*
[‡] *K.U.Leuven, FETEW, Belgium and University of Amsterdam, The Netherlands*
[§] *K.U.Leuven, Department of Mathematics, Belgium*
Email: marc.decamps@econ.kuleuven.ac.be, marc.goovaerts@econ.kuleuven.ac.be, Wim.Schoutens@wis.kuleuven.ac.be

Abstract

The Vasicek, the CIR and the CEV term structure models define the short rate process r as a linear diffusion with continuous scale $s'(r)$ and speed $m(r)$ densities. In this contribution, we permit the speed density to be discontinuous at the level r^*. Similarly to Gorovoi and Linetsky (2004), we obtain eigenfunction expansions for the prices of general contingent claims when analytical expressions exist for the continuous case. We interpret the resulting term structure as a continuous-time version of the Self Exciting Threshold AutoRegressive models (SETAR) popular in time series analysis. Finally, we calibrate a SET model with two Vasicek regimes to the U.S. yield curve.

1. INTRODUCTION

The dynamic of the short-term interest rate has received considerable attention in the financial literature. Among many others, the Vasicek (1977) and the Cox-Ingersoll-Ross (1985) models define the short rate as a linear diffusion with mean reverting instantaneous drift that guarantees the stationarity of the process. The Vasicek model assumes a constant instantaneous volatility while the volatility of the CIR model vanishes rapidly when the short rate falls off in order to make zero unattainable. The Vasicek and the CIR models are very tractable as closed-form expressions exist for the transition density and the bond price. Unfortunately, these models partly fail in capturing the empirical behavior of short rate time series.

The Japanese interest rates since the Asian crisis illustrate the unadequacy of classical models. As mentioned by Goldstein and Keirstead (1997) and Gorovoi and Linetsky (2004), the Japanese short-term rate during the period 1996 – 2003 remained at a very low level, but with a rather high volatility. The Vasicek model is consistent with high volatility at low interest rate regime but the probability for the short rate to become negative is not negligible whereas the CIR model precludes negative interest rate through a low volatility near zero. The second difficulty encountered when modeling the Japanese term structure of interest rates relates to the so-called Zero Interest Rate Policy (ZIRP). In February 1999, the Bank of Japan adopted the ZIRP by providing the necessary liquidity to offer very cheap credit against the deflationary pressure. The ZIRP was abandoned in Augustus 2000 and reactivated on March 19, 2001. The changes in the policy of the Bank of Japan have resulted in a regime switching behavior of the short-term rate depending whether the ZIRP is activated to maintain the short rate near zero or deactivated to permit short rate around 0.5 percent. Goldstein and Keirstead (1997) provide a solution to this problem by imposing a reflecting or an absorbing boundary to the short rate process while Black (1995) proposes the use of a shadow rate. As explained in details in Gorovoi and Linetsky (2004), analytical expressions can be recovered by using eigenfunction expansions for both models.

The U.S. interest rates have a similar regime switching feature depending on the level of the short rate. As mentioned in Pfann, Schotman and Tschering (1996), during the period 1979 – 1982 interest rates were very high and extremely volatile. They argue that the volatility of the U.S. interest rates plummets when the short rate falls below 8.5 percent. Markov switching regime models were introduced in the literature to capture this behavior. Under these models, the short rate switches between discrete regimes each of them driven by a diffusion process with distinct drift and volatility. Ait-Sahalia (1996) criticizes such models on their time-inhomogeneous feature and argues in favor of a short rate process with bimodal transition probability, both modes corresponding to a different regime. This can be achieved through a diffusion process with highly nonlinear instantaneous drift and volatility, see Ait-Sahalia (1996).

In this paper, we define the short-term rate as a linear diffusion on the state space $I = (e_1, e_2)$ and we allow the speed density to be discontinuous at the level $r^* \in (e_1, e_2)$. In case $s'(r)$ is continuous, the short rate $\{r_t\}_{t\geq 0}$ is solution of a stochastic differential equation with two regimes

$$dr_t = \begin{cases} \mu_1(r_t)dt + \sigma_1(r_t)dW_t, & e_1 < r_t \leq r^* \\ \mu_2(r_t)dt + \sigma_2(r_t)dW_t, & r^* < r_t < e_2, \end{cases}$$

where $\{W_t\}_{t\geq 0}$ is a standard Brownian motion under the risk-neutral measure[1], the differences $\mu_2(r^*) - \mu_1(r^*)$ and $\sigma_2(r^*) - \sigma_1(r^*)$ are finite. The resulting term structure is a time-continuous version of the Self Exciting Threshold AutoRegressive (SETAR) time series model used by Pfann, Schotman and Tschering (1996). We write for short that the process r is a SET diffusion with two regimes. Following Linetsky (2002) and Gorovoi and Linetsky (2004), we obtain closed-form expressions for the transition density and the prices of European-style contingent claims that facilitate the calibration of the models.

The paper is organized as follows. In section 2, we start with a description of the model. We define the short-term rate process as a diffusion with our specific assumptions on the scale and the speed densities. We introduce the notion of pricing semi-group and state-price density. In section 3, we adapt the results of Linetsky (2002) to our models and we obtain eigenfunction expansions when the spectrum of the pricing semi-group is discrete. We give also analytical results in terms of special functions for the case of two Vasicek regimes. Finally, we calibrate a SET model with two Vasicek regimes to the U.S. zero-yield curve.

[1] We assume in the sequel that the market risk premium is time-homogeneous and also discontinuous at the level r^*, thus the scale and the speed densities are discontinuous under both the historical and the risk-neutral measures.

2. THE MODEL

One-factor models of term structure are based on a single state variable which is usually the short-term rate. Most of the models define the short rate as a linear diffusion X with infinitesimal drift $\mu(x)$ and infinitesimal volatility $\sigma(x)$ taking values on an interval $I = (e_1, e_2)$. Let $\{\mathcal{P}_t\}_{t\geq 0}$ be the semi-group of operators such that for every bounded function f

$$\begin{aligned}(\mathcal{P}_t f)(x) &:= E_x[f(X_t)] \\ &= \int_I f(y)p(t;x,y)dy\end{aligned}$$

where $p(t;x,y)$ is the transition probability *w.r.t.* the Lebesgue measure. The scale $s'(x)$ and speed $m(x)$ densities are defined by

$$\begin{aligned}s'(x) &= \exp\left\{-\int^x \frac{2\mu(z)}{\sigma^2(z)}dz\right\} \\ m(x) &= \frac{2}{s'(x)\sigma^2(x)} \qquad (1)\end{aligned}$$

and give rise to the next representation of the infinitesimal generator of X, $\mathcal{G} : f \in \mathcal{D} \to g$:

$$\int_{[a,b)} g(x)m(dx) = \frac{df}{ds}(b) - \frac{df}{ds}(a); \quad \forall a, b \in I$$

where $\frac{df}{ds}(x) = \lim_{y\to x} \frac{f(y)-f(x)}{s(y)-s(x)}$ is the s-*derivative*, acting on the domain

$$\mathcal{D} = \left\{f : f, \mathcal{G}f \in C_b(I), \frac{df}{ds}(x) \text{ exists, conditions at } e_1 \text{ and } e_2\right\}.$$

Usual assumptions are the continuity of the functions μ, μ', σ, σ' and σ''. As mentioned in the introduction, we consider rather that μ and σ are discontinuous at the level r^* where $\mu(r^*_+) - \mu(r^*_-)$ and $\sigma(r^*_+) - \sigma(r^*_-)$ are finite. This implies that $m(x)$ and $s''(x)$ can be discontinuous.

The price of a contingent claim with payout $h \in C_b(I)$ is the expectation under some risk neutral measure of the discounted payments. Gorovoi and Linetsky (2004) introduce the pricing semi-group $\{\hat{\mathcal{P}}_t\}_{t\geq 0}$

$$\begin{aligned}(\hat{\mathcal{P}}_t h)(x) &:= E_x\left[e^{-\int_0^t X_s ds} h(X_t)\right] \\ &= \int_I \hat{p}(t;x,y)h(y)dy\end{aligned}$$

where $\hat{p}(t;x,y)$ is called the state-price density and can be interpreted as the prices of fundamental securities, or Arrow-Debreu securities that yield 1 only if the short rate equals y at time to maturity. We can replicate any European-style contingent claims with continuous payout $c(x,t)$ and final payoff $h(x)$ by purchasing a portfolio of these basic securities and determine the price as

$$\int_I \int_0^t \hat{p}(\tau;x,y)c(y,\tau)dyd\tau + \int_I \hat{p}(t;x,y)h(y)dy, \qquad (2)$$

see e.g. Beaglehole and Tenney (1991).

When X takes non-negative values, $\{\hat{\mathcal{P}}_t\}_{t\geq 0}$ is the semi-group of a linear diffusion killed at a rate x. Let $\hat{X}$ be the non-conservative linear diffusion with scale and speed densities defined by (1) sent to a cemetery ∂ when the additive functional $\int_0^t X_s ds$ exceeds an independent exponential random variable τ with parameter 1, then

$$\begin{aligned}(\hat{\mathcal{P}}_t h)(x) &= E_x[h(X_t)1_{(\tau<t)}] \\ &= E_x[h(\hat{X}_t)]\end{aligned}$$

assuming that $f(\partial)=0$. The generator of $\hat{X}$ is defined by $\hat{\mathcal{G}} : f \in \mathcal{D} \to g$

$$\int_{[a,b)} g(x)m(dx) = \frac{df}{ds}(b) - \frac{df}{ds}(a) - \int_{[a,b)} f(x)k(x)dx$$

where $k(x) = m(x)xdx$ is the killing measure and acts on the same domain as $\mathcal{G}$. We refer to Gorovoi and Linetsky (2004) and Linetsky (2002) for more details on pricing semi-group.

3. EIGENFUNCTION EXPANSIONS

The Vasicek and the CIR models are very popular since the transition probability and the state-price density are known in closed form. When analytical solutions exist for both regimes 1 and 2, we can use the spectral theory to recover tractable expressions. According to Ito and McKean (1974), the transition density *w.r.t.* to the Lebesgue measure associated to the semi-group with infinitesimal generator $\mathcal{G}$ satisfies the partial differential equation

$$\mathcal{G}u(t;x) = \frac{d}{dt}u(t;x),$$

and can be constructed by means of an eigenfunction expansion. The eigenfunctions $\varphi_\lambda(x)$ are the continuous solutions with continuous scale derivative $\frac{d\varphi_\lambda}{ds}(x)$ of the Sturm-Liouville problem

$$-(\mathcal{G}u)(x) = \lambda u(x), \qquad \forall x \in I = (e_1, e_2) \tag{3}$$

for some $\lambda \in \mathcal{C}$ such that $\varphi_\lambda(x)$ is m-square integrable and satisfies appropriate boundary conditions. The ordinary differential equation (3) can be solved as soon as analytical solutions exist on the intervals $(e_1, r^*]$ and $[r^*, e_2)$.

Applications of spectral theory to finance are recent, we refer to the pioneer papers of Linetsky (2002) and of Gorovoi and Linetsky (2004). For most of the models in finance, the spectrum of the state-price density is a countable sequence $\{\lambda_n\}_{n\in\mathcal{N}}$ and the spectral decomposition reduces to the series

$$\hat{p}(t;x,y) = m(y)\sum_{n=0}^{+\infty} e^{-\lambda_n t}\varphi_n(x)\varphi_n(y) \tag{4}$$

where $\varphi_n(x)$ is the *normalized* eigenfunction associated to λ_n and $m(y)$ is the speed density. As mentioned in Gorovoi and Linetsky (2004), if the short-term rate can reach negative value, the pricing semi-group is not the contraction semi-group of a linear diffusion. In this case, the eigenvalues λ_n are no longer guaranteed to be positive. This can lead to economical contradiction as the yield of the zero-coupon bonds converges to λ_0 for increasing maturities. In this section, we adapt the Proposition 3.3 in Gorovoi and Linetsky (2004) to the present situation. The following theorem gives a method to obtain the eigenfunctions and the eigenvalues for the transition probability and the state-price density of SET models with discrete spectra. We give also analytical results in terms of special functions for the case of two Vasicek regimes.

Theorem 3.1 *Assume the linear diffusion with infinitesimal volatility* $\sigma(r) = \sigma_1(r)1_{(r<r^*)} + \sigma_2(r)1_{(r\geq r^*)}$*, infinitesimal drift* $\mu(r) = \mu_1(r)1_{(r<r^*)} + \mu_2(r)1_{(r\geq r^*)}$ *and with domain* $I = (e_1, e_2)$*;* $-\infty \leq e_1 < e_2 \leq +\infty$*. Let* $\phi_\lambda(r)$ *be the unique (to some multiplicative constant) continuous solution with continuous scale derivative* $\frac{d\phi_\lambda}{ds}(r)$ *of the ODE*

$$-\frac{1}{2}\sigma_1^2(r)u''(r) - \mu_1(r)u'(r) + ru(r) = \lambda u(r) \qquad \forall r \in (e_1, r^*]$$

such that $\int_{e_1}^{r^*} |\phi_\lambda(r)|^2 m(r)dr < +\infty$ *and* $\phi_\lambda(r)$ *satisfies the appropriate condition at* e_1*.*

Let $\psi_\lambda(r)$ *be the unique continuous solution with continuous scale derivative* $\frac{d\psi_\lambda}{ds}(r)$ *of the ODE*

$$-\frac{1}{2}\sigma_2^2(r)u''(r) - \mu_2(r)u'(r) + ru(r) = \lambda u(r) \qquad \forall r \in [r^*, e_2)$$

such that $\int_{r^*}^{e_2} |\psi_\lambda(r)|^2 m(r)dr < +\infty$ *and* $\psi_\lambda(r)$ *satisfies the appropriate condition at* e_2*.*

Then, the eigenvalues $\lambda_0 < \lambda_1 < \lambda_2 < \cdots$ *of the Sturm-Liouville problem (3) associated to the state-price density are the zeros of the Wronskian* $\omega(\lambda)$

$$\omega(\lambda) := \phi_\lambda(r^*)\frac{\psi_\lambda}{ds}(r^*) - \psi_\lambda(r^*)\frac{\phi_\lambda}{ds}(r^*) = 0$$

and the eigenfunctions $\varphi_n(r)$ *read*

$$\varphi_n(r) = \begin{cases} \sqrt{\dfrac{\psi_{\lambda_n}(r^*)}{\Delta_n \phi_{\lambda_n}(r^*)}}\phi_{\lambda_n}(r), & e_1 < r \leq r^* \\ \sqrt{\dfrac{\phi_{\lambda_n}(r^*)}{\Delta_n \psi_{\lambda_n}(r^*)}}\psi_{\lambda_n}(r), & r^* \leq r < e_2, \end{cases}$$

where $\Delta_n = \frac{d\omega(\lambda)}{d\lambda}|_{\lambda=\lambda_n}$*.*

Proof. The proof is similar to Proposition 3.3 in Gorovoi and Linetsky (2004). Following Lemma 1 in Linetsky (2002), there exists a unique (to some multiplicative constant) solution $\phi_{\lambda_n}(r)$ to the *ODE*

$$-\frac{1}{2}\sigma_1^2(r)u''(r) - \mu_1(r)u'(r) + ru(r) = \lambda_n u(r)$$

on the interval $I = (e_1, e_2)$ which is m-square integrable in a neighborhood of e_1 and satisfies the appropriate condition at the boundary e_1. As the eigenfunction $\varphi_n(r)$ is also m-square integrable in a neighborhood of e_1 and satisfies the appropriate condition at the boundary e_1, $\varphi_n(r)$ must be equal to $\phi_{\lambda_n}(r)$ up to a constant. Similarly, we can deduce that $\varphi_n(r)$ is also equal to $\psi_{\lambda_n}(r)$ up to a constant. We conclude that there exists a non-zero constant A_n such that $\phi_{\lambda_n}(r) = A_n\psi_{\lambda_n}(r)$. The Wronskian is defined as

$$\omega(\lambda) := \phi_\lambda(r)\frac{\psi_\lambda}{ds}(r) - \psi_\lambda(r)\frac{\phi_\lambda}{ds}(r),$$

it is easy to check that $\omega(\lambda)$ depends only on λ as $\phi_\lambda(r)$ and $\psi_\lambda(r)$ are both continuous solutions of $-(\mathcal{G}u) = \lambda u$. Moreover, $\phi_{\lambda_n}(r) = A_n\psi_{\lambda_n}(r)$ implies that $w(\lambda) = 0$ for $\lambda = \lambda_n$. From Theorem 5 in Linetsky (2002), we know that $||\phi_{\lambda_n}(r)|| = A_n\omega'(\lambda_n)$ and thus, $\varphi_n(r)$ is continuous at r^* and $||\varphi_n(r)|| = 1$. □

3.1. SET Vasicek model

The Vasicek model (1977) defines the short rate as the Gaussian process solution of the stochastic differential equation

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t,$$

with state space $I = (-\infty, +\infty)$. Similarly, the Self Exciting Threshold Vasicek model is driven by the short rate process solution of

$$dr_t = \begin{cases} \kappa_1(\theta_1 - r_t)dt + \sigma_1 dW_t, & -\infty < r_t \le r^* \\ \kappa_2(\theta_2 - r_t)dt + \sigma_2 dW_t, & r^* \le r_t < +\infty. \end{cases}$$

The resulting process is a scalar diffusion with scale density

$$s'(r) = \begin{cases} e^{\frac{k_1(\theta_1 - r)^2}{\sigma_1^2}}, & -\infty < r \le r^* \\ e^{\frac{k_1(\theta_1 - r)^2}{\sigma_1^2} - \frac{k_1(\theta_1 - r^*)^2}{\sigma_1^2} + \frac{k_2(\theta_2 - r^*)^2}{\sigma_2^2}}, & r^* \le r < +\infty; \end{cases}$$

and speed density $m(r) = 2/(s'(r)\sigma^2(r))$ with $\sigma^2(r) = \sigma_1^2 1_{(r<r^*)} + \sigma_2^2 1_{(r\ge r^*)}$ discontinuous at the level r^*. A direct application of Theorem 3.1 leads to the next Proposition.

Proposition 3.2 *The functions $\phi_\lambda(r)$ and $\psi_\lambda(r)$ defined in Theorem 3.1 corresponding to the SET Vasicek model are given by*

$$\begin{aligned} \phi_\lambda(r) &= e^{z_1^2/4} D_{\mu_1}(-(\alpha_1 - z_1)) \\ \psi_\lambda(r) &= e^{z_2^2/4} D_{\mu_2}(\alpha_2 - z_2) \end{aligned}$$

where $z_1 = \frac{\sqrt{2\kappa_1}}{\sigma_1}(\theta_1 - r)$ and $z_2 = \frac{\sqrt{2\kappa_2}}{\sigma_2}(\theta_2 - r)$; $\alpha_1 = \sigma_1^2\sqrt{2/\kappa_1^3}$ and $\alpha_2 = \sigma_2^2\sqrt{2/\kappa_2^3}$; and $D_\mu(z)$ are the parabolic cylinder functions of parameters $\mu_1 = \sigma_1^2/2\kappa_1^3 + (\lambda_1 - \theta_1)/\kappa_1$ and $\mu_2 = \sigma_2^2/2\kappa_2^3 + (\lambda_2 - \theta_2)/\kappa_2$.

Proof. The function $\phi_\lambda(r)$ of Theorem 3.1 is solution of the *ODE*

$$-\frac{1}{2}\sigma_1^2 u'' + \kappa_1(\theta_1 - r)u' + ru = \lambda u, \quad r \in (-\infty, r*]. \tag{5}$$

We look for solutions in the form $u(r) = e^{z_1^2/4} v(z_1)$ with $z_1 = \frac{\sqrt{2\kappa_1}}{\sigma_1}(\theta_1 - r)$. Substituting $u(r)$ in equation (5), we obtain that $v(z)$ satisfies the Weber-Hermite equation

$$v'' + \left(\frac{1}{2} + \mu_1 - \frac{(\alpha_1 - z^2)}{4}\right) v = 0, \quad z \in \left[\sqrt{2\kappa_1}\theta_1/\sigma_1, +\infty\right)$$

with $\mu_1 = \sigma_1^2/2\kappa_1^3 + (\lambda_1 - \theta_1)/\kappa_1$ and $\alpha_1 = \sigma_1^2\sqrt{2/\kappa_1^3}$. The solution m-square integrable in a neighborhood of $+\infty$ is the parabolic cylinder function $D_{\mu_1}(-(\alpha_1 - z_1))$. With similar arguments, we find that $\psi_\lambda(r) = e^{z_2^2/4} D_{\mu_2}(\alpha_2 - z_2)$ is the unique solution of

$$-\frac{1}{2}\sigma_2^2 u'' + \kappa_2(\theta_2 - r)u' + ru = \lambda u, r \in [r*, +\infty).$$

that is m-square integrable in a neighborhood of $+\infty$. □

4. CALIBRATION TO THE U.S. ZERO-YIELD CURVE

In this section, we calibrate a SET Vasicek model to the U.S. bond market. The data set consists of 15 STRIPS bond prices obtained from Datastream on 14/12/2003. We minimize the root squared error between the STRIPS yield curve and the model yield curve[2]. The optimization procedure provides the parameter estimates $\kappa_1 = 0.3999$, $\theta_1 = 0.0606$ and $\sigma_1 = 0.0105$; $\kappa_2 = 0.197$, $\theta_2 = 0.097$ and $\sigma_2 = 0.0284$; $r^* = 0.0813$ for the SET Vasicek and $\kappa = 0.2563$, $\theta = 0.0654$ and $\sigma = 5.119e^{-5}$ for the Vasicek model. Figure 1 compares the STRIPS yield curve with the Vasicek and the SET Vasicek yield curves (with $k = 120$ terms). The SET vasicek model improves significantly the fit to the current term structure. The volatility estimate of the Vasicek model is almost zero which is consistent with low levels of the U.S. short-term rate rate but in contradiction with higher regimes. Finally, we draw the same conclusions than Pfann, Schotman and Tschering (1996), the U.S. short-term rate have two distinct regimes with a discontinuity of the volatility around 8.5 percent.

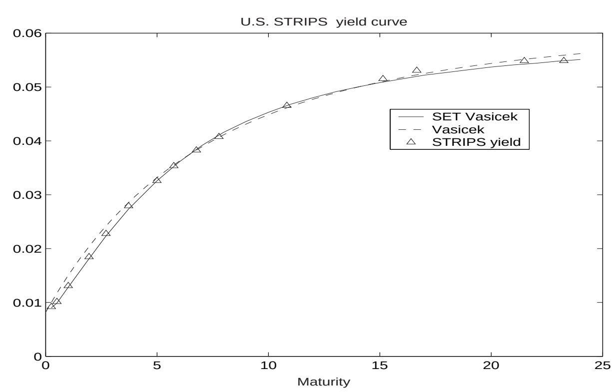


Figure 1: U.S. zero-yield curve on 14/12/2003, Datastream

[2]We only consider bonds with maturities larger than 2Y to avoid numerical problems.

References

[1] Ait-Sahalia Y. (1996): Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, **9**, 385-426.

[2] Ait-Sahalia Y. (1999): Transition probabilities for interest rates and other non-linear diffusions. *Journal of finance*, **54**(4), 1361-1395.

[3] Ait-Sahalia Y. (2002): Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, **70**, No. 1 (January, 2002), 223-262.

[4] Beaglehole D.R. and Tenney M.S. (1991): General solutions of some interest-rate claim pricing equation. *Journal of fixed income*, **2**, 69-83.

[5] Black, F. (1995): Interest rate as options. *Journal of Finance*, **50**, 1371-1376.

[6] Cox J.C., Ingersoll J.E. and Ross S.A. (1985): A Theory of the Term Structure of Interest Rate. *Econometrica*, **53**, 385-407.

[7] Gorovoi V. and Linetsky V. (2004): Black's model of interest rates as options, eigenfunction expansions and Japanese interest rate. *Mathematical finance*, **14**(1), 49-78.

[8] Linetsky V. (2002): The spectral decomposition of the option value. *Working Paper*, Northwestern University.

[9] Itô K. and McKean H.P. (1974): *Diffusion processes and their Sample Paths*. Springer, Berlin.

[10] Pfann G.A., Schotman P.C. and Tschering R. (1996): Nonlinear interest rate dynamics and implications for the term structure. *Journal of Econometrics*, **74**, 149-176.

[11] Vasicek O.A. (1977): An equilibrium characterization of the term structure. *Journal of Financial Economics*, **5**, 177-188.

COMONOTONIC APPROXIMATIONS FOR OPTIMAL PORTFOLIO SELECTION PROBLEMS: THE CASE OF TERMINAL WEALTH

Jan Dhaene[†], Steven Vanduffel[§], Marc J. Goovaerts[†], Rob Kaas[‡] and David Vyncke[§]

[†] *KULeuven, Belgium and University of Amsterdam, The Netherlands*
[‡] *University of Amsterdam, The Netherlands*
[§] *KULeuven, Belgium*
Email: `Steven.Vanduffel@econ.kuleuven.ac.be`

Abstract

We investigate multiperiod portfolio selection problems in a Black & Scholes type market where a basket of 1 riskless and m risky securities are traded continuously. We look for the optimal allocation of wealth within the class of 'constant mix' portfolios. First, we consider the portfolio selection problem of a decision maker who invests money at predetermined points in time in order to obtain a target capital at the end of the time period under consideration. Several optimality criteria and their interpretation within Yaari's dual theory of choice under risk are presented. We propose accurate approximations based on the concept of comonotonicity, as studied in Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002a,b). Our analytical approach avoids simulation, and hence reduces the computing effort drastically. This paper is a reduced version of Dhaene, Vanduffel, Goovaerts, Kaas and Vyncke (2004).

1. INTRODUCTION

Strategic portfolio selection is the process used to identify the best allocation of wealth among a basket of securities for an investor with a given consumption/saving behavior over a given investment horizon. The basket of available securities will typically be a selection of risky assets such as stocks, bonds and real estate, and risk-free components such as cash and money market instruments. The individual investor or the asset manager chooses an initial asset mix and a particular tactical trading strategy within a given set of strategies, according to which he will buy and sell risky and risk-free assets, during the whole time period under consideration.

In this paper we will investigate multi-period optimal portfolio selection problems in a Black & Scholes (1973) lognormal setting. We will assume that the investor has to choose the optimal investment strategy for a given consumption or savings pattern, within the class of constant mix strategies. In this paper we will consider only the terminal wealth problem. Similar results can be obtained for the so-called reserving problem.

In the *terminal wealth problem*, the decision maker will invest a given series of positive saving amounts $\alpha_0, \alpha_{1,} \ldots, \alpha_{n-1}$ at predetermined times $0, 1, \ldots, n-1$ such that his terminal wealth at time n will reach or exceed some target capital K with a sufficiently large probability.

As terminal wealth is a sum of dependent lognormal random variables, its distribution function cannot be determined exactly and is too cumbersome to work with. Therefore, we will present accurate approximations for the distribution function at hand. The first approximation that we will consider for the distribution of terminal wealth will be called the ‘comonotonic upper bound’ as it is an upper bound for the exact distribution in the convex order sense. It is derived by keeping the marginal distributions exact but approximating the copula that describes the dependency structure between the random accumulation factors involved by the comonotonic copula.

Our second approximation for the exact distribution is based on the technique of conditioning. In this approach, the marginal distributions are changed and as a result the copula describing the dependency structure is replaced by the comonotonic copula. We will call this the ‘comonotonic lower bound’ approach as it can be proven that it is a lower bound in the sense of convex ordering. Especially this lower bound will perform very accurately as an approximation to the exact distribution.

The approximations that we propose have several advantages. First, for any given investment strategy they provide an accurate and easy to compute approximation for any risk measure that is additive for comonotonic risks, such as distortion risk measures (VaR and TailVaR for instance). Second, it turns out that for the comonotonic approximations we propose, the optimal investment mix can be found on the mean-variance efficient frontier. Third, the comonotonic approximations reduce the multivariate randomness of the multiperiod problem to a univariate randomness.

The proposed methodology can be used to solve several personal finance problems: for instance the so-called ‘saving for retirement problem’. In this case, one wants to retire in n years with a ‘nest egg’ of K — in real terms, i.e. in today’s Euro’s. How much does one have to save monthly — in real terms — in order to assure a $(1-\epsilon)$ chance to reach the retirement financial goal? Clearly the answer will depend on the investment mix. The theory on comonotonicity gives a quick, elegant and accurate answer to this question.

As the time horizon that we consider is long (typically 10, 20 or more years), assuming a Gaussian model seems to be appropriate, at least approximately, by the Central Limit Theorem. In order to verify whether the theoretical setup can be approximately compared with the data generating mechanism of real situations, we refer to Cesari & Cremonini (2003). For the period 1997-1999, the authors conclude that weekly (and longer period) returns can be considered as normal and independent. Daily returns on the other hand are both non-normal and autocorrelated.

The paper is organized as follows. In Sections 2 and 3 we present some results concerning risk measures, comonotonicity, the Black & Scholes setting, constant mix portfolios and mean-variance analysis that will be used throughout the paper. Next, the problem of finding optimal investment strategies in a general multivariate final wealth model with savings at discrete points in time is analyzed in Section 4. To the best of our knowledge, determining optimal investment strategies for terminal wealth problems by means of the comonotonic approach, as presented in Section 4, is new. We refer the interested reader to Dhaene, Vanduffel, Goovaerts, Kaas & Vyncke (2004) for a more extended and complete version of this paper.

2. RISK MEASURES AND COMONOTONICITY

In this section, we will introduce some definitions and present some results related to risk measures and comonotonicity that will be used throughout this paper. More details about comonotonicity can be found in Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002a,b), more details about the relation between risk measures and comonotonicity can be found in Dhaene, Vanduffel, Tang, Goovaerts, Kaas & Vyncke (2004).

2.1. Risk measures

A risk measure summarizes the information contained in the distribution function of a random variable in one single real number. For a random variable X, the p-quantile risk measure is defined as

$$Q_p\left[X\right] = \inf\left\{x \in \mathbb{R} \mid F_X(x) \geq p\right\}, \qquad p \in (0,1),$$

where $F_X(x) = \Pr\left[X \leq x\right]$ and by convention, $\inf\{\phi\} = +\infty$. A related risk measure is denoted by $Q_p^+\left[X\right]$ and is defined by

$$Q_p^+\left[X\right] = \sup\left\{x \in \mathbb{R} \mid F_X(x) \leq p\right\}, \qquad p \in (0,1),$$

where by convention $\sup\{\phi\} = -\infty$. If F_X is strictly increasing, both risk measures will coincide for all values of p. In this case, we can also define the $(1-p)$-quantiles by

$$Q_{1-p}\left[X\right] = \sup\left\{x \in \mathbb{R} \mid \overline{F}_X(x) \geq p\right\}, \qquad p \in (0,1),$$

where $\overline{F}_X(x) = 1 - F_X(x)$.

In the sequel, we will always consider random variables with finite mean. The Conditional Tail Expectation (CTE) at level p will be denoted by $\text{CTE}_p\left[X\right]$. It is defined by

$$CTE_p\left[X\right] = E\left[X \mid X > Q_p\left[X\right]\right], \qquad p \in (0,1).$$

The CTE measures the right tail of the distribution function. We will also need a risk measure that measures the left tail of the distribution function. Therefore, we introduce the Conditional Left Tail Expectation, which is defined by

$$CLTE_p\left[X\right] \; = E\left[X \mid X < Q_p^+\left[X\right]\right].$$

One can prove that the following relation holds between CTE and $CLTE$:

$$CLTE_{1-p}\left[X\right] = -CTE_p\left[-X\right], \qquad p \in (0,1).$$

2.2. Comonotonic bounds for sums of dependent random variables

A random vector $\underline{Y} = (Y_0, Y_1, \ldots, Y_n)$ is said to be comonotonic if

$$(Y_0, Y_1, \ldots, Y_n) \stackrel{d}{=} (F_{Y_0}^{-1}(U), F_{Y_1}^{-1}(U), \ldots, F_{Y_n}^{-1}(U)),$$

where U is a random variable which is uniformly distributed on the unit interval and where the notation $\stackrel{d}{=}$ stands for 'equality in distribution'.
For any random vector $\underline{X} = (X_0, X_1, \ldots, X_n)$, we will call its comonotonic counterpart any random vector with the same marginal distributions and with the comonotonic dependency structure. The comonotonic counterpart of $\underline{X} = (X_0, X_1, \ldots, X_n)$ will be denoted by $\underline{X}^c = (X_0^c, X_1^c, \ldots, X_n^c)$. Hence for any random vector $\underline{X} = (X_0, X_1, \ldots, X_n)$, we have

$$(X_0^c, X_1^c, \ldots, X_n^c) \stackrel{d}{=} (F_{X_0}^{-1}(U), F_{X_1}^{-1}(U), \ldots, F_{X_n}^{-1}(U)).$$

It can be proven that a random vector is comonotonic if and only if all its marginals are non-decreasing functions (or all are non-increasing functions) of the same random variable.

The random variable X is said to precede the random variable Y in the stop-loss order sense, notation $X \leq_{sl} Y$, if X has lower stop-loss premiums than Y:

$$\mathrm{E}[(X-d)_+] \leq \mathrm{E}[(Y-d)_+], \qquad -\infty < d < +\infty. \tag{1}$$

On the other hand, X is said to precede Y in the convex order sense, notation $X \leq_{cx} Y$, if $X \leq_{sl} Y$ and in addition $\mathrm{E}[X] = \mathrm{E}[Y]$. In Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002a) a proof for the following theorem can be found.

Theorem 2.1 (Convex bounds for sums of random variables)
For any random vector $(X_0, X_1, \ldots, X_n)$ *and any random variable* Λ*, we have that*

$$\sum_{i=0}^{n} E\left[X_i \mid \Lambda\right] \leq_{cx} \sum_{i=0}^{n} X_i \leq_{cx} \sum_{i=0}^{n} F_{X_i}^{-1}(U). \tag{2}$$

The theorem above states that the least attractive random vector $(X_0, X_1, \ldots, X_n)$ with given marginals F_{X_i}, in the sense that the sum of its components is largest in the convex order, has the *comonotonic* joint distribution, which means that it has the joint distribution of the random vector $\left(F_{X_0}^{-1}(U), F_{X_1}^{-1}(U), \ldots, F_{X_n}^{-1}(U)\right)$.
The random variable $S^c = \sum_{i=0}^{n} F_{X_i}^{-1}(U)$ will be called the comonotonic upper bound of $S = \sum_{i=0}^{n} X_i$, whereas the random variable $S^l = \sum_{i=0}^{n} E\left[X_i \mid \Lambda\right]$ will be referred to as a lower bound for S.
The random vector $\left(E\left[X_0 \mid \Lambda\right], E\left[X_1 \mid \Lambda\right], \ldots, E\left[X_n \mid \Lambda\right]\right)$ will in general not have the same marginal distributions as $(X_0, X_1, \ldots, X_n)$. If one can find a conditioning random variable Λ with the property that all random variables $E\left[X_i \mid \Lambda\right]$ are non-increasing functions of Λ (or all are non-decreasing functions of Λ), then the lower bound $S^l = \sum_{i=0}^{n} E\left[X_i \mid \Lambda\right]$ is a sum of n comonotonic random variables. The advantage of the comonotonic dependency structure is that any distortion risk measure of a sum of comonotonic random variables equals the sum of the risk measures of the marginals involved. For the quantile risk measures defined above, we find for all $p \in (0,1)$:

$$Q_p\left[S^c\right] = \sum_{i=0}^{n} Q_p\left[X_i\right],$$

$$Q_p^+\left[S^c\right] = \sum_{i=0}^{n} Q_p^+\left[X_i\right].$$

For the CTE and the CLTE a similar result can be proven, provided all marginal distributions F_{X_i} are continuous:

$$CTE_p\left[S^c\right] = \sum_{i=0}^{n} CTE_p\left[X_i\right], \text{ provided all } F_{X_i} \text{ are continuous,}$$
$$CLTE_p\left[S^c\right] = \sum_{i=0}^{n} CLTE_p\left[X_i\right], \text{ provided all } F_{X_i} \text{ are continuous.}$$

2.3. Sums of lognormal random variables

Consider the sum

$$S = \sum_{i=0}^{n} \alpha_i \, e^{Z_i}$$

where the α_i are non-negative constants and the Z_i are linear combinations of the components of the random vector $(Y_1,\ Y_2, \ldots,\ Y_n)$ which is assumed to have a multivariate normal distribution:

$$Z_i = \sum_{j=1}^{n} \lambda_{ij} \, Y_j.$$

Let U be uniformly distributed on the unit interval. The comonotonic upper bound $S^c = \sum_{i=0}^{n} F^{-1}_{\alpha_i e^{Z_i}}(U)$ of S is given by

$$S^c = \sum_{i=0}^{n} \alpha_i \, e^{E[Z_i] + \sigma_{Z_i} \, \Phi^{-1}(U)}.$$

Taking into account the additivity property, the following expressions can be derived for the risk measures associated with S^c:

$$Q_p\left[S^c\right] = Q_p^+\left[S^c\right] = \sum_{i=0}^{n} \alpha_i \, e^{E[Z_i] + \sigma_{Z_i} \, \Phi^{-1}(p)},$$
$$CTE_p\left[S^c\right] = \sum_{i=0}^{n} \alpha_i \, e^{E[Z_i] + \frac{1}{2}\sigma^2_{Z_i}} \frac{\Phi\left(\sigma_{Z_i} - \Phi^{-1}(p)\right)}{1-p},$$
$$CLTE_p\left[S^c\right] = \sum_{i=0}^{n} \alpha_i \, e^{E[Z_i] + \frac{1}{2}\sigma^2_{Z_i}} \frac{1 - \Phi\left(\sigma_{Z_i} - \Phi^{-1}(p)\right)}{p}, \qquad p \in (0,1).$$

In order to define a comonotonic lower bound S^l for S, we choose a conditioning random variable Λ which is a linear combination of the Y_j:

$$\Lambda = \sum_{j=1}^{n} \beta_j \, Y_j.$$

After some computations, we find that the lower bound $S^l = \sum_{i=0}^{n} \alpha_i\, E\left[e^{Z_i} \mid \Lambda\right]$ is given by

$$S^l = \sum_{i=0}^{n} \alpha_i\, e^{E[Z_i]+\frac{1}{2}\left(1-r_i^2\right)\sigma_{Z_i}^2 + r_i\ \sigma_{Z_i}\ \Phi^{-1}(U)},$$

where the uniformly distributed random variable U follows from $\Phi^{-1}(U) \equiv \frac{\Lambda - E(\Lambda)}{\sigma_\Lambda}$, and r_i is the correlation between Z_i and Λ.

If all r_i are positive, then S^l is a comonotonic sum. Hence, assuming that all r_i are positive, we find the following expressions for the risk measures associated with S^l:

$$\begin{aligned}
Q_p\left[S^l\right] = Q_p^+\left[S^l\right] &= \sum_{i=0}^{n} \alpha_i\, e^{E[Z_i]+\frac{1}{2}\left(1-r_i^2\right)\sigma_{Z_i}^2 + r_i\, \sigma_{Z_i}\ \Phi^{-1}(p)}, \\
CTE_p\left[S^l\right] &= \sum_{i=0}^{n} \alpha_i\, e^{E[Z_i]+\frac{1}{2}\sigma_{Z_i}^2}\, \frac{\Phi\left(r_i\, \sigma_{Z_i} - \Phi^{-1}(p)\right)}{1-p}, \\
CLTE_p\left[S^l\right] &= \sum_{i=0}^{n} \alpha_i\, e^{E[Z_i]+\frac{1}{2}\sigma_{Z_i}^2}\, \frac{1-\Phi\left(r_i\, \sigma_{Z_i} - \Phi^{-1}(p)\right)}{p}, \qquad p \in (0,1)\,.
\end{aligned}$$

Several examples in Dhaene, Denuit, Goovaerts, Kaas & Vyncke (2002b) show that especially the lower bound approximation performs very well as an approximation for the risk measures for sums of lognormals. Therefore, in the sequel we will only study how the lower bound enables us to approximate efficiently "optimal portfolio's".

3. STOCHASTIC RETURN PROCESSES

3.1. The Black & Scholes setting

Throughout the paper, we will assume the classical continuous-time framework that was pioneered by Merton (1971) and is nowadays mostly referred to as the Black & Scholes (1973) setting. We suppose that there is a market in which $(m+1)$ securities (assets or investment accounts) are traded continuously. One of the assets is the risk free asset. Let $P^0(0) = P^0 > 0$ be the current price, at time 0, of 1 unit of the risk free asset, whereas $P^0(t)$ is its price at time t. This price is assumed to evolve according to the following ordinary differential equation:

$$\frac{\mathrm{d}P^0(t)}{P^0(t)} = r\mathrm{d}t,$$

with $r > 0$. On the other hand, let $P^i(0) = P^i > 0$ be the current price, at time 0, of 1 unit of risky asset i, whereas $P^i(t)$ is the price at time t (including reinvestment of dividend income) of one unit of risky asset i. The price process $P^i(t)$ evolves according to a geometric Brownian motion stochastic process, represented by the following stochastic differential equation:

$$\frac{\mathrm{d}P^i(t)}{P^i(t)} = \mu_i \mathrm{d}t + \sum_{j=1}^{d} \overline{\sigma}_{ij}\, \mathrm{d}W^j(t), \qquad i = 1, \ldots, m, \tag{3}$$

where $\left(W^1(\tau),\ W^2(\tau),\ldots,\ W^d(\tau)\right)$ is a d-dimensional standard Brownian motion process. The $W^i(\tau)$ are mutually independent standard Brownian motions.

The m-dimensional vector $\boldsymbol{\mu}^T = (\mu_1 \ \cdots \ \mu_m)$ is called the drift vector of the risky assets. We will assume that $\boldsymbol{\mu} \neq r\mathbf{1}$, with $\mathbf{1}^T = (1\ 1 \cdots 1)$.

The $(m \times d)$ matrix $\overline{\boldsymbol{\Sigma}}$ defined by

$$\overline{\boldsymbol{\Sigma}} = \begin{pmatrix} \overline{\sigma}_{11} & \overline{\sigma}_{12} & \cdots & \overline{\sigma}_{1d} \\ \overline{\sigma}_{21} & \overline{\sigma}_{22} & \cdots & \overline{\sigma}_{2d} \\ \cdots & \cdots & \cdots & \cdots \\ \overline{\sigma}_{m1} & \overline{\sigma}_{m2} & \cdots & \overline{\sigma}_{md} \end{pmatrix}$$

is called the diffusion matrix. Further, we define the $(m \times m)$ matrix $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \overline{\boldsymbol{\Sigma}} \cdot \overline{\boldsymbol{\Sigma}}^T = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_m^2 \end{pmatrix}, \tag{4}$$

with coefficients σ_{ij} and σ_i^2 given by $\sigma_{ij} = \sum_{k=1}^{d} \overline{\sigma}_{ik}\,\overline{\sigma}_{jk}$ and $\sigma_i^2 = \sigma_{ii}$. We have that $\sigma_{ij} = \sigma_{ji}$. The matrix $\boldsymbol{\Sigma}$ is called the variance-covariance matrix. We will assume that $\boldsymbol{\Sigma}$ is positive definite. In particular, this assumption implies that all σ_i are strictly positive. Hence, all m risky assets are indeed risky. It also implies that $\boldsymbol{\Sigma}$ is non-singular, meaning that its determinant is strictly positive, and hence $\boldsymbol{\Sigma}$ has a matrix inverse $\boldsymbol{\Sigma}^{-1}$. As we will see further on, the elements of the matrix $\boldsymbol{\Sigma}$ describe the covariances between the yearly returns of the different investment accounts.

We define the process $B^i(\tau)$ by

$$B^i(\tau) = \frac{1}{\sigma_i} \sum_{j=1}^{d} \overline{\sigma}_{ij} W^j(\tau).$$

Rewriting equation (3), we find

$$\frac{\mathrm{d}P^i(t)}{P^i(t)} = \mu_i \mathrm{d}t + \sigma_i\, \mathrm{d}B^i(t), \qquad i = 1, \ldots, m. \tag{5}$$

The solution to equation (5) is

$$P^i(t) = P^i\ \exp\left\{ \left(\mu_i - \frac{1}{2}\sigma_i^2 \right) t + \sigma_i\ B^i(t) \right\},$$

which means that $\frac{P^i(t)}{P^i}$ is lognormally distributed with parameters $\left(\mu_i - \frac{1}{2}\sigma_i^2\right) t$ and $\sigma_i^2\, t$, respectively. This implies that the expectation and standard deviation of the price of asset i at time t are given by

$$E\left[P^i(t)\right] = P^i\, e^{\mu_i t},$$
$$\sigma\left[P^i(t)\right] = P^i\, e^{\mu_i t} \sqrt{e^{\sigma_i^2\, t} - 1}.$$

Let $k = 1, 2, \ldots$. Investing an amount of 1 at time $k-1$ in investment account i will grow to the random amount $e^{Y_k^i}$ at time k, where Y_k^i denotes the yearly return in year k of account i. One finds that

$$Y_k^i = \left(\mu_i - \frac{1}{2}\sigma_i^2\right) + \sigma_i \left(B^i(k) - B^i(k-1)\right).$$

Hence, it follows that the random yearly returns Y_k^i of asset i are independent and have identical normal distributions with

$$\begin{aligned} E\left[Y_k^i\right] &= \mu_i - \frac{1}{2}\sigma_i^2, \\ Var\left[Y_k^i\right] &= \sigma_i^2, \\ Cov\left[Y_k^i,\ Y_l^j\right] &= \begin{cases} 0 & k \neq l \\ \sigma_{ij} & k = l \end{cases}. \end{aligned}$$

As announced earlier, the matrix $\mathbf{\Sigma}$ is the variance-covariance matrix of the yearly return vector $(Y_k^1, Y_k^2, \ldots, Y_k^n)$.

3.2. Constant mix strategies

Assume one can invest wealth in one or more of the $m+1$ assets as defined above. Let $\boldsymbol{\pi}(t)^T = (\pi_1(t),\ \pi_2(t), \ldots,\ \pi_m(t))$ be the vector describing the portfolio process, i.e. $\pi_i(t)$ is the fraction of the wealth that is invested in risky asset i at time t. The residual, i.e. $1 - \sum_{i=1}^n \pi_i(t)$ is invested in the risk free asset, or, if negative, finances the risky asset purchases. A negative proportion invested in the risk free asset means borrowing (going short) on the risk free asset.

We will restrict to constant portfolios $\boldsymbol{\pi}(t)^T = \boldsymbol{\pi}^T = (\pi_1,\ \pi_2, \ldots,\ \pi_m)$, which means that the fractions invested in the different assets remain constant over time. Investing according to a constant portfolio process implies that one has to follow a dynamic trading strategy. Indeed, as the risky asset returns evolve randomly, one has to trade at each instant in order to keep the fractions invested in the different assets constant. Such investment strategies are known as constant mix strategies, or also as constant proportional investment strategies. Optimality of constant mix strategies in a utility theory setting is considered in Merton (1971).

Let us now consider one unit of a security that is constructed according to the continuously rebalanced investment strategy $(\pi_1,\ \pi_2, \ldots,\ \pi_m)$, and let $P(t)$ be the price of that unit at time t, with $P(0) = P$. One can prove that the price process $P(t)$ evolves according to the dynamics

$$\begin{aligned} \frac{\mathrm{d}P(t)}{P(t)} &= \sum_{i=1}^m \pi_i \frac{\mathrm{d}P^i(t)}{P^i(t)} + \left(1 - \sum_{i=1}^m \pi_i\right) \frac{\mathrm{d}P^0(t)}{P^0(t)} \\ &= \left(\sum_{i=1}^m \pi_i\ (\mu_i - r) + r\right) \mathrm{d}t + \sum_{i=1}^m \pi_i\, \sigma_i \mathrm{d}B^i(t). \end{aligned} \tag{6}$$

For a non-zero vector $\boldsymbol{\pi}$, One can verify that the process $B(\tau)$ defined by

$$B(\tau) = \frac{1}{\sqrt{\boldsymbol{\pi}^T \cdot \mathbf{\Sigma} \cdot \boldsymbol{\pi}}} \sum_{i=1}^m \pi_i\ \sigma_i\ B^i(\tau).$$

is a standard Brownian motion. Equation (6) can then be rewritten as

$$\frac{\mathrm{d}P(t)}{P(t)} = \mu(\boldsymbol{\pi})\,\mathrm{d}t + \sigma(\boldsymbol{\pi})\,\mathrm{d}B(t) \tag{7}$$

with

$$\mu(\boldsymbol{\pi}) = r + \boldsymbol{\pi}^T \cdot (\boldsymbol{\mu} - r\mathbf{1}) \quad \text{and} \quad \sigma^2(\boldsymbol{\pi}) = \boldsymbol{\pi}^T \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{\pi}, \tag{8}$$

where $\mathbf{1}$ is the m-vector $(1\,1\cdots 1)$. The solution to equation (7) is

$$P(t) = P\exp\left\{\left(\mu(\boldsymbol{\pi}) - \frac{1}{2}\sigma^2(\boldsymbol{\pi})\right)t + \sigma(\boldsymbol{\pi})\,B(t)\right\},$$

with expectation and standard deviation given by

$$E\left[P(t)\right] = P\,e^{\mu(\boldsymbol{\pi})t},$$
$$\sigma\left[P(t)\right] = P\,e^{\mu(\boldsymbol{\pi})t}\sqrt{e^{\sigma^2(\boldsymbol{\pi})t} - 1}.$$

The stochastic differential equation (7) was derived by Merton (1971, 1990), see also Rubinstein (1991).

Let k be a strictly positive integer. Investing according to investment strategy $\boldsymbol{\pi}$, an amount of 1 at time $k-1$ will grow to the random amount $e^{Y_k(\boldsymbol{\pi})}$ at time k, where $Y_k(\boldsymbol{\pi})$ denotes the yearly return in year k of investment strategy $\boldsymbol{\pi}$. One finds that

$$Y_k(\boldsymbol{\pi}) = \left(\mu(\boldsymbol{\pi}) - \frac{1}{2}\sigma^2(\boldsymbol{\pi})\right) + \sigma(\boldsymbol{\pi})\left(B(k) - B(k-1)\right).$$

Hence, the random yearly returns $Y_k(\boldsymbol{\pi})$ of the constantly rebalanced portfolio $\boldsymbol{\pi}$ are independent and identically distributed normal random variables with

$$E\left[Y_k(\boldsymbol{\pi})\right] = \mu(\boldsymbol{\pi}) - \frac{1}{2}\sigma^2(\boldsymbol{\pi}),$$
$$Var\left[Y_k(\boldsymbol{\pi})\right] = \sigma^2(\boldsymbol{\pi}).$$

The price $P(k)$ can be written in terms of the yearly returns as follows:

$$P(k) = P\,\exp\left(Y_1(\boldsymbol{\pi}) + Y_2(\boldsymbol{\pi}) + \cdots + Y_k(\boldsymbol{\pi})\right).$$

3.3. Markowitz mean-variance analysis

In 1990, Harry M. Markowitz received the Nobel Prize in Economics (shared with William F. Sharpe and Merton H. Miller) for his theory on portfolio selection under uncertainty. As mentioned in the press release of the Royal Swedish Academy of Sciences, Markowitz's theory can be considered as the first approach to solving the problem that each investor faces, namely how to find the optimal trade-off between risk and return, i.e. how to find the optimal investment strategy under the two conflicting objectives of high expected return versus low risk of the investment portfolio. Markowitz proposed a way to reduce the complicated and multidimensional problem of finding the optimal portfolio with respect to a large number of different assets to a conceptual simple two-dimensional problem, known as mean-variance analysis.

Several variants of the classical single-period mean-variance problem exist. Here, we will consider the formulation that we will need later on in the paper. Among all constant mix portfolios $\boldsymbol{\pi}$ with a given portfolio drift $\mu(\boldsymbol{\pi}) = \mu$, we look for the one with the smallest volatility $\sigma(\boldsymbol{\pi})$. Hence, for any given value of μ, we want to find the solution of the following problem:

$$\text{Min}_{\boldsymbol{\pi}}\ \sigma^2(\boldsymbol{\pi}) \ \text{ subject to } \mu(\boldsymbol{\pi}) = \mu. \tag{9}$$

We will denote the portfolio that corresponds to the minimum in (9) by $\boldsymbol{\pi}^{\mu}$.
The assumption that $\boldsymbol{\mu} \neq r\mathbf{1}$, together with the assumptions that the variance-covariance matrix is positive definite and that short-selling is allowed implies that there exists a unique local global minimum for problem (9). A Lagrange optimization yields:

$$\sigma^2(\boldsymbol{\pi}^{\mu}) = \frac{(\mu - r)^2}{(\boldsymbol{\mu} - r\mathbf{1})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{\mu} - r\mathbf{1})} \tag{10}$$

and

$$\boldsymbol{\pi}^{\mu} = (\mu - r)\,\frac{\boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{\mu} - r\mathbf{1})}{(\boldsymbol{\mu} - r\mathbf{1})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{\mu} - r\mathbf{1})}. \tag{11}$$

Note that $\sigma^2(\boldsymbol{\pi}^{\mu})$ and $\boldsymbol{\pi}^{\mu}$ are well-defined, because the inverse of a positive definite matrix is also positive definite.

The efficient frontier refers to the set of all solutions $\{(\sigma(\boldsymbol{\pi}^{\mu}), \mu)\}$ for the optimization problem (9). From (10) we see that the efficient frontier consists of two straight lines in the (σ, μ)-plane:

$$\begin{aligned} \mu &= r + \sqrt{(\boldsymbol{\mu} - r\mathbf{1})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{\mu} - r\mathbf{1})}\,\sigma(\boldsymbol{\pi}^{\mu}), \qquad \mu \geq r, \\ \mu &= r - \sqrt{(\boldsymbol{\mu} - r\mathbf{1})^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{\mu} - r\mathbf{1})}\,\sigma(\boldsymbol{\pi}^{\mu}), \qquad \mu \leq r. \end{aligned} \tag{12}$$

The portfolios $\boldsymbol{\pi}^{\mu}$ belonging to the efficient frontier are called mean-variance efficient portfolios. Portfolios on the lower branch are irrelevant from a mean-variance optimization viewpoint as they lead to a positive volatility while their drift is lower than r. The upper branch $\{(\sigma(\boldsymbol{\pi}^{\mu}), \mu) \mid \mu \geq r\}$ is referred to as the 'Capital Market Line'.

In the following, we will call portfolios $\boldsymbol{\pi}$ that fulfill the condition $\mathbf{1}^T \times \boldsymbol{\pi} = 1$ risky-assets-only portfolios because such portfolios consist only of risky assets. It can be proven that if we only consider risky-assets-only portfolios, the efficient frontier is a hyperbola in the mean - standard deviation space (provided there are at least two risky assets with different drift). Now consider the risky-assets-only global minimal variance portfolio $\boldsymbol{\pi}^{(m)}$, i.e. the portfolio that is the solution of the following problem:

$$\text{Min}_{\boldsymbol{\pi}}\ \sigma^2(\boldsymbol{\pi}) \ \text{ subject to } \mathbf{1}^T \cdot \boldsymbol{\pi} = 1.$$

This portfolio and its drift are given by

$$\begin{aligned} \boldsymbol{\pi}^{(m)} &= \frac{\boldsymbol{\Sigma}^{-1} \cdot \mathbf{1}}{\mathbf{1}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \mathbf{1}}, \\ \mu\left(\boldsymbol{\pi}^{(m)}\right) &= \frac{\mathbf{1}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \boldsymbol{\mu}}{\mathbf{1}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot \mathbf{1}}. \end{aligned}$$

One can prove that under the condition

$$\mu\left(\boldsymbol{\pi}^{(m)}\right) > r,$$

the Capital Market Line is at a tangent to the upper branch of the hyperbola that corresponds to the efficient frontier of risky-asset-only portfolios. When $\mu\left(\boldsymbol{\pi}^{(m)}\right) < r$, the decreasing part of the efficient frontier (12) will be tangent to the lower branch of the hyperbola.
Let us now assume that $\mu\left(\boldsymbol{\pi}^{(m)}\right) \neq r$. The portfolio that corresponds to the point of intersection between the efficient frontier (12) and the risky-assets-only efficient frontier is called the 'tangency portfolio', and is denoted by $\boldsymbol{\pi}^{(t)}$. The assumption that $\mu\left(\boldsymbol{\pi}^{(m)}\right) \neq r$ implies that $\mu\left(\boldsymbol{\pi}^{(t)}\right) \neq r$. One can easily verify that $\boldsymbol{\pi}^{(t)}$ is given by

$$\boldsymbol{\pi}^{(t)} = \frac{\boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{\mu} - r\mathbf{1})}{\mathbf{1}^T \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{\mu} - r\mathbf{1})}. \tag{13}$$

Note that (11) can be rewritten as

$$\boldsymbol{\pi}^{\mu} = \left(\frac{\mu - r}{\mu(\boldsymbol{\pi}^{\mathbf{t}}) - r}\right)\boldsymbol{\pi}^{(t)}.$$

This means that every mean-variance efficient portfolio $\boldsymbol{\pi}^{\mu}$ consists of a fraction $(\frac{\mu-r}{\mu(\boldsymbol{\pi}^{(\mathbf{t})})-r})$ invested in the risky-assets-only portfolio $\boldsymbol{\pi}^{(t)}$ and a fraction $(1 - \frac{\mu-r}{\mu(\boldsymbol{\pi}^{(\mathbf{t})})-r})$ invested in the risk free asset. Mean-variance optimizing investors only differ in terms of which fraction of their wealth they put in the tangency portfolio.
The result that all mean-variance investors will hold only two kinds of portfolios (or mutual funds), the exclusively risky portfolio $\boldsymbol{\pi}^{(t)}$ and the risk free asset, is often called a Mutual Fund Theorem or a Two Fund Separation Theorem.

In case $\mu\left(\boldsymbol{\pi}^{(m)}\right) > r$ is fulfilled, also the inequality $\mu\left(\boldsymbol{\pi}^{(t)}\right) > r$ holds. The Capital Market Line can then be rewritten as

$$\mu = r + \left(\frac{\mu\left(\boldsymbol{\pi}^{(\mathbf{t})}\right) - r}{\sigma\left(\boldsymbol{\pi}^{(t)}\right)}\right)\sigma\left(\boldsymbol{\pi}^{\mu}\right).$$

This equation describes the drift of the return for an investor as related to the volatility that he is willing to accept. The slope $\frac{\mu\left(\boldsymbol{\pi}^{(\mathbf{t})}\right)-r}{\sigma(\boldsymbol{\pi}^{t})}$ is referred to as the 'Sharpe ratio'. It can be interpreted as the price of risk reduction: It shows by how much the drift increases if the volatility increases by 1 unit.

4. SAVING AND TERMINAL WEALTH

4.1. General problem description

In this section, we will consider the problem of how to invest periodic saving amounts in order to reach some target capital at a predetermined future time n. Let α_i be the positive amount that will be invested at time i, ($i = 0, 1, 2, \ldots, n$). We assume that these amounts are invested according to a constant mix portfolio $\boldsymbol{\pi}$ as defined in Section 3.2. The choice of the constant portfolio mix has to be made at time 0. An amount of 1 unit invested at time i will grow to the random amount $e^{\sum_{j=i+1}^{n} Y_j(\boldsymbol{\pi})}$ at time n.

Let $W_j(\boldsymbol{\pi})$ be the wealth at time j, defined by the following recursive relation:

$$W_j(\boldsymbol{\pi}) = W_{j-1}(\boldsymbol{\pi})\; e^{Y_j(\boldsymbol{\pi})} + \alpha_j, \qquad j = 1, \ldots, n, \tag{14}$$

with initial value $W_0(\boldsymbol{\pi}) = \alpha_0$. Hence, $W_j(\boldsymbol{\pi})$ is the wealth that will be available at time j, including the savings amount α_j at time j. The realization of $W_j(\boldsymbol{\pi})$ will be known at time j, and depends on the investment returns (stochastic part) and on the savings (deterministic part) in the past. Note that the random variables $Y_j(\boldsymbol{\pi})$ are i.i.d. and normal distributed with parameters $\mu(\boldsymbol{\pi})$ and $\sigma(\boldsymbol{\pi})$ as defined in (8).

From the recursion (14) for the wealth process, we find the following explicit expression for terminal wealth $W_n(\boldsymbol{\pi})$:

$$W_n(\boldsymbol{\pi}) = \sum_{i=0}^{n} \alpha_i\; e^{\sum_{j=i+1}^{n} Y_j(\boldsymbol{\pi})}. \tag{15}$$

By convention, $\sum_{i=m}^{n} b_i$ is set equal to 0 if $m > n$.

Within the expected utility theory framework of Von Neumann & Morgenstern (1947), the investor could choose the investment strategy $\boldsymbol{\pi}$ that maximizes his expected utility of final wealth:

$$\max_{\boldsymbol{\pi}} E\left[u(W_n(\boldsymbol{\pi}))\right],$$

where u is the utility function he uses to appreciate the different levels of final wealth.

Another approach, within the framework of Yaari's (1987) dual theory of choice under risk, is to choose the optimal investment strategy as the one that maximizes the distorted expectation of final wealth:

$$\max_{\boldsymbol{\pi}} \rho_f\left[W_n(\boldsymbol{\pi})\right], \tag{16}$$

where f is the investor's distortion function and ρ_f is the 'distorted expectation', determined with $f\left(\Pr\left(W_n(\boldsymbol{\pi}) > x\right)\right)$:

$$\rho_f\left[W_n(\boldsymbol{\pi})\right] = -\left(\int_{-\infty}^{0} 1 - f\left(\Pr\left(W_n(\boldsymbol{\pi}) > x\right)\right)\right) \mathrm{d}x + \int_{0}^{\infty} f\left(\Pr\left(W_n(\boldsymbol{\pi}) > x\right)\right) \mathrm{d}x.$$

While in utility theory, choosing among risks is performed by comparing expected values of transformed wealth levels (utilities), in Yaari's theory the quantities that are compared are the 'distorted expectations' of wealth levels. The distorted expectation of final wealth $W_n(\boldsymbol{\pi})$ can be interpreted as an expectation of $W_n(\boldsymbol{\pi})$ evaluated with a 'distorted probability measure' in the sense of a Choquet-integral, see Denneberg (1994). The decision maker acts in order to maximize the distorted expectation of final wealth.

For a distortion function f_p, $0 < p < 1$, given by

$$f_p(x) = \begin{cases} 0 & 0 \le x < p \\ 1 & p \le x \le 1 \end{cases} \tag{17}$$

we find

$$\begin{aligned}\rho_{f_p}\left[W_n\left(\boldsymbol{\pi}\right)\right] &= Q^{+}_{1-p}\left[W_n\left(\boldsymbol{\pi}\right)\right] \\ &= \sup\left\{x\in\mathbb{R} \mid \Pr\left(W_n\left(\boldsymbol{\pi}\right)>x\right)\geq p\right\}.\end{aligned} \tag{18}$$

The optimization problem (16) with distortion function given by (17) determines the optimal investment strategy as the one that maximizes the largest amount that will be reached with a probability of at least p.

For the convex distortion function g_p, $0<p<1$, given by

$$g_p(x)=\begin{cases} 0 & 0\leq x<p \\ \dfrac{x-p}{1-p} & p\leq x\leq 1 \end{cases} \tag{19}$$

we find

$$\rho_{g_p}\left[W_n\left(\boldsymbol{\pi}\right)\right]=CLTE_{1-p}\left[W_n\left(\boldsymbol{\pi}\right)\right].$$

In Yaari's theory, a decision maker is called risk-averse if he has a convex distortion function. Hence, the optimization problem (16) with distortion function (19) can be interpreted as the problem to be solved by a risk-averse decision maker with distortion function g_p. The optimal investment strategy is the one that maximizes the conditional expected value of final wealth, given that the p-target capital is not reached.

For a more detailed comparison between the two theories of choice under risk and their relation to risk measures, see e.g. Dhaene, Vanduffel, Tang, Goovaerts, Kaas & Vyncke (2004).

4.2. Comonotonic lower bound approximations

From (15), we see that $W_n\left(\boldsymbol{\pi}\right)$ is a sum of non-independent lognormal random variables. As it is impossible to determine the distribution function of $W_n\left(\boldsymbol{\pi}\right)$ analytically, we will derive a convex order lower bound $W_n^l\left(\boldsymbol{\pi}\right)$ for $W_n\left(\boldsymbol{\pi}\right)$.
Rewriting $W_n\left(\boldsymbol{\pi}\right)$ as

$$W_n\left(\boldsymbol{\pi}\right)=\sum_{i=0}^{n}\alpha_i\, e^{Z_i},$$

we see that we can apply the results of Section 2.3 with

$$\begin{aligned} Z_i &= Y_{i+1}\left(\boldsymbol{\pi}\right)+Y_{i+2}\left(\boldsymbol{\pi}\right)+\cdots+Y_n\left(\boldsymbol{\pi}\right), \\ E\left[Z_i\right] &= (n-i)\left[\mu\left(\boldsymbol{\pi}\right)-\frac{1}{2}\sigma^2\left(\boldsymbol{\pi}\right)\right], \\ \sigma^2_{Z_i} &= (n-i)\,\sigma^2\left(\boldsymbol{\pi}\right). \end{aligned}$$

In order to define a convex lower bound $W_n^l\left(\boldsymbol{\pi}\right)$ for $W_n\left(\boldsymbol{\pi}\right)$, we choose the conditioning random variable as follows:

$$\Lambda\left(\boldsymbol{\pi}\right)=\sum_{j=1}^{n}\beta_j\left(\boldsymbol{\pi}\right)Y_j\left(\boldsymbol{\pi}\right)$$

where the coefficients $\beta_j(\boldsymbol{\pi})$ are as follows:

$$\beta_j(\boldsymbol{\pi}) = \sum_{k=0}^{j-1} \alpha_k e^{-k\mu(\boldsymbol{\pi})}$$

It follows that for this choice of the parameters $\beta_j(\boldsymbol{\pi})$, the variance of the lower bound will be close to the variance of $W_n(\boldsymbol{\pi})$, provided $\sigma^2(\boldsymbol{\pi})$ is small enough.
From Section 2.3, we find

$$W_n^l(\boldsymbol{\pi}) = \sum_{i=0}^{n} \alpha_i e^{(n-i)\mu(\boldsymbol{\pi}) - \frac{1}{2} r_i^2(\boldsymbol{\pi})(n-i)\sigma^2(\boldsymbol{\pi}) + r_i(\boldsymbol{\pi})\sqrt{n-i}\sigma(\boldsymbol{\pi})\Phi^{-1}(U)} \tag{20}$$

where the coefficients $r_i(\boldsymbol{\pi})$ are given by

$$r_i(\boldsymbol{\pi}) = \frac{\sum_{j=i+1}^{n} \sum_{k=0}^{j-1} \alpha_k e^{-k\mu(\boldsymbol{\pi})}}{\sqrt{n-i}\sqrt{\sum_{j=1}^{n}\left(\sum_{k=0}^{j-1} \alpha_k e^{-k\mu(\boldsymbol{\pi})}\right)^2}}.$$

Note that the correlation coefficients $r_i(\boldsymbol{\pi})$ are non-negative which implies that $W^l(\boldsymbol{\pi})$ is a comonotonic sum of lognormal random variables.
The following expression can be derived for the risk measure $Q_{1-p}^{+}(W_n^l(\boldsymbol{\pi}))$, $p \in (0,1)$:

$$Q_{1-p}^{+}\left[W_n^l(\boldsymbol{\pi})\right] = Q_{1-p}\left[W_n^l(\boldsymbol{\pi})\right] = \sum_{i=0}^{n} \alpha_i e^{(n-i)\left(\mu(\boldsymbol{\pi}) - \frac{1}{2} r_i^2(\boldsymbol{\pi})\sigma^2(\boldsymbol{\pi})\right) - r_i(\boldsymbol{\pi})\sqrt{n-i}\sigma(\boldsymbol{\pi})\Phi^{-1}(p)}, \tag{21}$$

while for $CLTE_{1-p}\left[W_n^l(\boldsymbol{\pi})\right]$ we find

$$CLTE_{1-p}\left[W_n^l(\boldsymbol{\pi})\right] = \sum_{i=0}^{n} \alpha_i \, e^{(n-i)\mu(\boldsymbol{\pi})} \, \frac{1 - \Phi\left(r_i(\boldsymbol{\pi})\sqrt{n-i}\,\sigma(\boldsymbol{\pi}) + \Phi^{-1}(p)\right)}{1-p}.$$

4.3. Determining the investment strategy that maximizes the target capital, for a given probability level

4.3.1. The p-target capital

For a given probability level $\frac{1}{2} < p < 1$ and a given investment strategy $\boldsymbol{\pi}$, we define the p-target capital K as the $(1-p)$-th order "$+$"-quantile of terminal wealth:

$$K = Q_{1-p}^{+}\left[W_n(\boldsymbol{\pi})\right]. \tag{22}$$

One immediately finds that

$$K = \sup\left\{x \in \mathbb{R} \mid \Pr\left[W_n(\boldsymbol{\pi}) > x\right] \geq p\right\}.$$

Hence, the target capital at probability level p can be interpreted as the maximal amount that will be available at time n, with a probability of at least p.

Now assume that a probability level p is fixed and that the optimal investment strategy $\boldsymbol{\pi}^*$ is determined as the one that maximizes the p-target capital. Denoting the optimal target capital by K^*, we have

$$K^* = \max_{\boldsymbol{\pi}} Q^+_{1-p} \left[W_n \left(\boldsymbol{\pi} \right) \right]. \tag{23}$$

Note that from (16) and (18), it follows that this optimization problem can be interpreted in terms of Yaari's dual theory of choice under risk.
Solving (23) is from a computational point of view a complicated problem because of the multi-dimensionality involved. Indeed, a 'time-dimensionality' occurs because $W_n(\boldsymbol{\pi})$ is a sum of n dependent accumulation factors. There is also a 'portfolio-dimensionality' involved as the maximum has to be determined over all portfolios $\boldsymbol{\pi}$. In the following section we will show how to get rid of this 'curse of dimensionality'.

4.3.2. The comonotonic lower bound for $W_n(\boldsymbol{\pi})$

We also propose to approximate the optimal investment strategy $\boldsymbol{\pi}^*$ by $\boldsymbol{\pi}^l$, where $\boldsymbol{\pi}^l$ is the investment strategy that maximizes $Q^+_{1-p}(W^l_n(\boldsymbol{\pi}))$. The p-target capital K^* is then approximated by K^l, which is given by

$$K^l = \max_{\boldsymbol{\pi}} Q^+_{1-p} \left[W^l_n \left(\boldsymbol{\pi} \right) \right].$$

It follows from (21) that for a given value of $\mu(\boldsymbol{\pi})$, the correlation coefficient is fixed and the quantile $Q_p(W^l_n(\boldsymbol{\pi}))$ is a decreasing function of $\sigma(\boldsymbol{\pi})$. Hence, $\boldsymbol{\pi}^l$ is an element of the set of efficient portfolios. The general maximization problem can be reduced to the following maximization problem:

$$K^l = \max_{\mu} Q^+_{1-p} \left[W^l_n \left(\boldsymbol{\pi}^{\mu} \right) \right]. \tag{24}$$

The approximated optimization problem (24) solves the curse of dimensionality. The multi-dimensionality caused by time n is reduced to one dimension by introducing the comonotonic dependency structure. Also the portfolio-dimensionality m is reduced to one dimension because the optimal solutions are to be found on the efficient frontier.

4.3.3. Constant savings amounts

In this subsection, we consider the special case that the saving amounts are constant. For each investment strategy $\boldsymbol{\pi}$ we look for the required periodic saving amount α that leads to a p-target capital equal to 1. From (22) we find that this saving amount α is given by

$$\alpha(\boldsymbol{\pi}) = \frac{1}{Q^+_{1-p}\left[\overline{W}_n(\boldsymbol{\pi})\right]},$$

with $\overline{W}_n(\boldsymbol{\pi})$ given by

$$\overline{W}_n(\boldsymbol{\pi}) = \sum_{i=0}^{n} e^{Y_{i+1}(\boldsymbol{\pi}) + Y_{i+2}(\boldsymbol{\pi}) + \cdots + Y_n(\boldsymbol{\pi})}.$$

The optimal investment strategy is now defined as the one that minimizes the period savings. Denoting the minimal saving amount by α^*, we have

$$\alpha^* = \min_{\boldsymbol{\pi}} \alpha\left(\boldsymbol{\pi}\right).$$

Note that in the case of constant saving amounts, the investment strategy that maximizes the p-target capital K for given saving amounts α is identical to the investment strategy that minimizes the periodic savings α for a given target capital K.

Now, we approximate $\overline{W}_n\left(\boldsymbol{\pi}\right)$ by $\overline{W}_n^l\left(\boldsymbol{\pi}\right)$ as explained in (20). The minimal periodic savings amount α^* is then approximated by α^l which is given by

$$\alpha^l = \min_{\mu} \frac{1}{Q_{1-p}^{+}\left[\overline{W}_n^l\left(\boldsymbol{\pi}^{\mu}\right)\right]}.$$

4.3.4. Numerical illustration

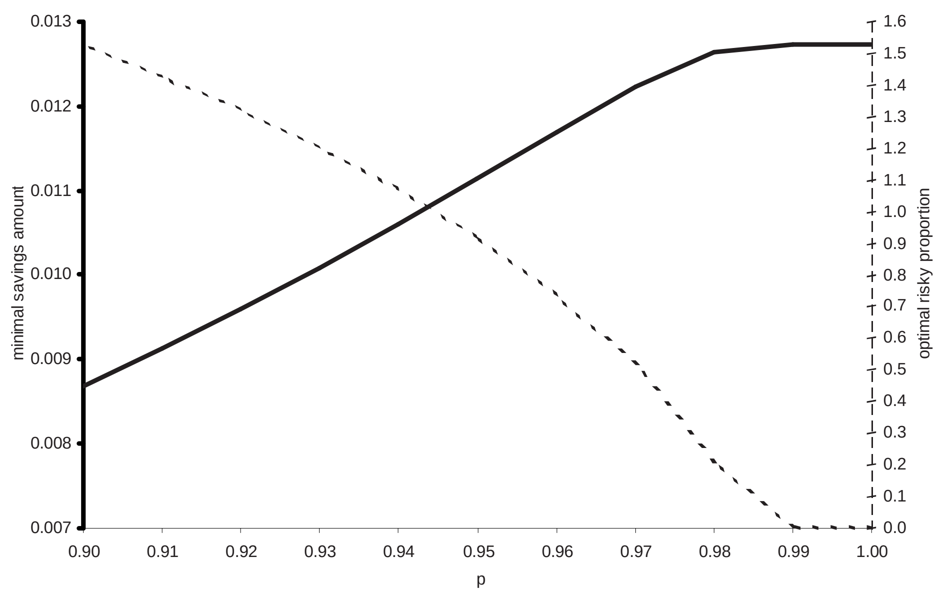


Figure 1: The minimal savings amount α^l (solid line - left scale) and the optimal risky proportion π^l (dashed line - right scale) as a function of p.

Consider a Black & Scholes market with a risk free asset with a yearly return $r = 0.03$ and two risky assets with yearly drifts equal to $\mu_1 = 0.06$ and $\mu_2 = 0.10$ respectively. The volatilities of the risky assets are given by $\sigma_1 = 0.10$ and $\sigma_2 = 0.20$. Pearson's correlation coefficient $\frac{\sigma_{12}}{\sigma_1\sigma_2}$ is given by 0.5. From (13) we find that the tangency portfolio is given by $\boldsymbol{\pi}^{(t)} = \left(\frac{5}{9}, \frac{4}{9}\right)$ with drift $\mu\left(\boldsymbol{\pi}^{(t)}\right) = 7/90$ and volatility $\sigma(\boldsymbol{\pi}^{(t)}) = \sqrt{\frac{43}{2700}}$.

We assume constant saving amounts α at times $0, 1, \ldots, 39$ and a target capital equal to 1 to be reached at time 40. In Figure 1, we consider the investment strategy that minimizes the yearly savings amount for different probability levels p of the target capital. The computations were performed with the lower bound approximation $\overline{W}^{l}_{40}(\boldsymbol{\pi})$ for $\overline{W}_{40}(\boldsymbol{\pi})$.
The solid line represents the (approximated) minimal savings amount α^l for different probability levels p of a target capital equal to 1 (left scale). As we see from the figure, increasing the required probability of reaching the target of 1, increases the optimal savings amount. Note that the required savings amount in case of the risk free investment, i.e. the one that corresponds to $p = 1$, is given by 0.0127.
The dashed line represents the (approximated) optimal risky proportion π^l to be invested in the tangency portfolio, for different probability levels p (right scale). As could be expected, increasing the probability of reaching the target capital decreases the optimal risky proportion in the portfolio.

Acknowledgements

Jan Dhaene, Marc Goovaerts, Steven Vanduffel and David Vyncke acknowledge the financial support of the Onderzoeksfonds K.U. Leuven (GOA/02: Actuariële, financiële en statistische aspecten van afhankelijkheden in verzekerings- en financiële portefeuilles).

References

[1] Björk T. (1998): *Arbitrage Theory in Continuous Time*. Oxford University Press, New York.

[2] Black F. and Scholes M. (1973): The pricing of options and corporate liabilities. *Journal of Political Economy*, **81** (May-June), 637-659.

[3] Cesari R. and Cremonini D. (2003): Benchmarking, portfolio insurance and technical analysis: a Monte Carlo comparison of dynamic strategies of asset allocation. *Journal of Economic Dynamics and Control*, **27**, 987-1011.

[4] Dhaene J., Denuit M., Goovaerts M.J., Kaas R.and Vyncke D. (2002a): The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics & Economics*, **31**(1), 3-33.

[5] Dhaene J., Denuit M., Goovaerts M.J., Kaas R. and Vyncke D. (2002b): The concept of comonotonicity in actuarial science and finance: applications. *Insurance: Mathematics & Economics*, **31**(2), 133-161.

[6] Dhaene J., Vanduffel S., Tang Q., Goovaerts M.J., Kaas R. and Vyncke D. (2004): Solvency capital, risk measures and comonotonicity: a review. *Working paper*, `www.kuleuven.ac.be/insurance, publications`.

[7] Dhaene J., Vanduffel S., Goovaerts M.J., Kaas R. and Vyncke D. (2004): Comonotonic approximations for optimal portfolio selection problems. *Working paper*, `www.kuleuven.ac.be/insurance, publications`.

[8] Emmer S., Klüppelberg C. and Korn R. (2001): Optimal portfolios with bounded capital at risk. *Mathematical Finance*, **11**(4), 365-384.

[9] Markowitz H. (1952): Portfolio selection. *Journal of Finance*, **7**, 77-91.

[10] Merton R. (1971): Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, **3**, 373-413.

[11] Milevsky M. (2003): Space-time diversification: Which dimension is better? *Journal of Risk*, **5**(2), to be published.

[12] Rubinstein M. (1991): Continuously rebalanced investment strategies. *The Journal of Portfolio Management*, Fall, 78-81.

[13] Samuelson P.A. (1989): The judgement of economic science on rational portfolio management: Timing and long horizon effects. *The Journal of Portfolio Management*, **16**(1), 4-12.

[14] Vanduffel S., Dhaene J., Goovaerts M. and Kaas R. (2003): The hurdle-race problem. *Insurance: Mathematics & Economics*, **33**, 405-413.

[15] Von Neumann J. and Morgenstern O. (1947): *Theory of games and economic behavior*. Princeton University Press, Princeton.

[16] Yaari M.E. (1987): The dual theory of choice under risk, *Econometrica*, **55**, 95-115.

APPROXIMATIONS OF THE DISTRIBUTION OF GENERAL ANNUITIES IN THE CASE OF TRUNCATE STOCHASTIC INTEREST RATES

Inge Koch and Ann De Schepper

University of Antwerp, Faculty of Applied Economics, Prinsstraat 13, B-2000 Antwerp
Email: `inge.koch@ua.ac.be`

Abstract

An important part of the current financial and actuarial research deals with the investigation of present value functions in the case of a stochastic interest rate. In the present contribution, it is shown how interest rates can be restricted to meet special types of financial or actuarial constraints. Approximate but analytical expressions are given for the distribution of different types of annuities, and their accuracy is illustrated graphically.

1. INTRODUCTION

Many of the problems in the current financial and actuarial research can be reduced to the problem of finding the distribution of the present value of a cash-flow in the form

$$V(t) = \sum_{i=1}^{n} \alpha(t_i)\, e^{-X(t_i)}, \tag{1}$$

where $0 < t_1 < t_2 < \cdots < t_n = t$, where $\alpha(t_i)$ is a (positive or negative) payment at time t_i, and where $X = \{X(t)\}$ is a stochastic process with $X(t_i)$ denoting the compounded rate of return for the period $[0, t_i]$.

There exists a broad range of stochastic processes that seem to be useful to model the stochastic interest rates, which is shown by the long list of papers investigating these models. However, in many cases, the model would be more realistic if the interest rates are not completely free, but restricted to some range of acceptable values. If for example the interest rates appearing in the cash-flow are nominal interest rates, they can not become negative. If an insurance contract guarantees a minimal return, the interest rate model should be adapted in order to meet this warranty. Due to special regulations, it can also be necessary to impose an upper limit for the yield of a financial effect.

In this paper, we want to introduce a model that meets these last requirements. We show how to adapt common models to these restrictions, and we show the influence on the present value of classical actuarial functions such as annuities. Except for some special cases (concerning the restrictions and concerning the actual stochastic model), as a consequence of the adaptation, the exact distribution of the present value can no longer be calculated analytically. Therefore, we will make use of an approximation by means of convex bounds, as introduced by Goovaerts et al. [4], and generalized in Dhaene et al. [2, 3].

2. METHODOLOGY

2.1. Restrictions on the interest rates

A. NON-NEGATIVE INTEREST RATES

A first and common restriction, needed in many financial applications, goes back to the fact that (if nominal rates are used) negative interest rates should be avoided. A possible solution to this problem can be reached by multiplying the compounded rate of return with the Heaviside-function $\mathcal{U}$, defined by

$$\mathcal{U}(x) = \begin{cases} 1 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0, \end{cases}$$

such that the discount factor in the present value becomes $e^{-X(t)\,\mathcal{U}(X(t))}$. See also [1]. With this adjustment, the compounded interest rate is kept equal to zero as long as the value of $X(t)$ is negative.

B. TRUNCATE INTEREST RATES WITH FIXED FLOOR AND CEILING

A more general solution consists of a truncate interest rate, by defining a ceiling and a floor for the interest rate – with the previous restriction as a special case. This can be done by mapping $X(t)$ on $c \in \mathbb{R}$ whenever X(t) exceeds c, and by mapping $X(t)$ on $f \in \mathbb{R}$ whenever $X(t)$ is smaller than f.

Definition 2.1 *Let $f, c \in \mathbb{R}$ with $f < c$. The truncate function $\mathcal{S}_f^c : \mathbb{R} \to [f, c]$ then is defined by*

$$\mathcal{S}_f^c(x) = \begin{cases} f & \textit{if} \;\; x < f \\ x & \textit{if} \;\; f \leq x \leq c \\ c & \textit{if} \;\; c < x. \end{cases}$$

The left plot of figure 1 shows a possible realisation of such a truncate interest rate. With this truncate function applied on the stochastic interest rate, the discount factor in the present value becomes $e^{-\mathcal{S}_f^c(X(t))}$.

C. TRUNCATE INTEREST RATES WITH LINEAR FLOOR AND CEILING

Since the stochastic variable $X(t)$ corresponds to the *cumulative* interest rate for the period $[0, t]$, it seems more appropriate to use a fixed floor and ceiling per unit time period, or a linear floor and ceiling for the whole time period. This results in the following alternative definition for the truncation of the interest rates.

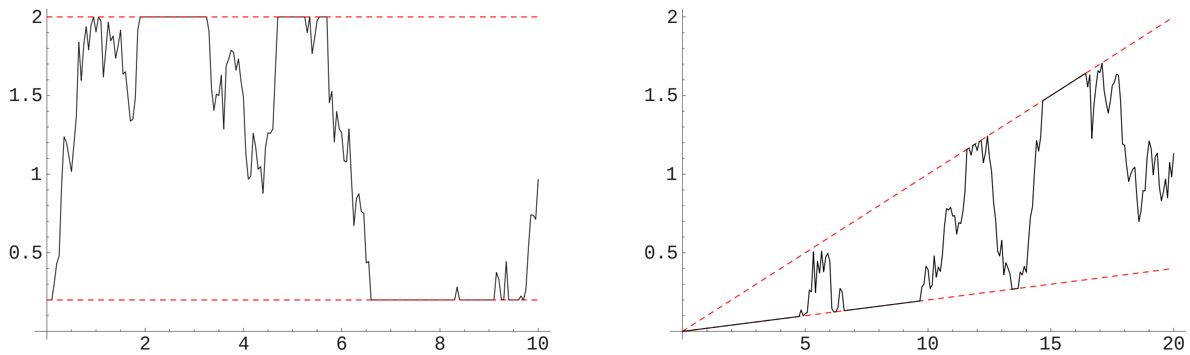


Figure 1: Example of a stochastic truncate (left) and a stochastic linear truncate (right) interest rate.

Definition 2.2 *Let $f, c \in \mathbb{R}$ with $f < c$. The linear truncate function $\tilde{\mathcal{S}}_f^c : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ is defined by*

$$\tilde{\mathcal{S}}_f^c(t,x) = \begin{cases} f \cdot t & \textit{if} \quad x < f \cdot t \\ x & \textit{if} \quad f \cdot t \leq x \leq c \cdot t \\ c \cdot t & \textit{if} \quad c \cdot t < x. \end{cases}$$

The right plot of figure 1 shows a possible realisation of such a linear truncate interest rate. With this linear truncate function applied on the stochastic interest rate, the discount factor in the present value becomes $e^{-\tilde{\mathcal{S}}_f^c(t,X(t))}$.

2.2. Convex bounds

Since the compounded rates of return $X(t_i)$ for successive periods only differ for the last part of the period, the present value of (1) is made up as a sum of rather dependent terms. As a consequence, it is nearly impossible to derive an exact analytical expression for the distribution of such a present value. In order to solve this problem, Goovaerts et al. [4] and Dhaene et al. [2] present bounds in convexity order. Following their method, the original sum $V(t)$ is replaced by a new sum, for which the components have the same marginal distributions as the components in the original sum, but with the most "dangerous" dependence structure that is possible, and for which the calculation of the distribution is much more easy.

In this subsection, we just briefly recall definitions and most important results about this approximation method. For details, we refer to Dhaene et al. [2].

Definition 2.3 *Let X and Y be two random variables, then X is said to be* smaller *than Y in* convex order sense, *(notation $X \leq_{cx} Y$), if and only if*

$$E[v(X)] \leq E[v(Y)]$$

for all real convex functions $v : \mathbb{R} \to \mathbb{R}$, provided the expectations exist.

In fact this ordering means that the variable Y is more likely to reach extreme values than it is the case for X, or, that the variable Y is more dangerous than X. Note that for such variables it is true that $E[X] = E[Y]$ and $Var[X] \leq Var[Y]$.

Theorem 2.1 *Let $X_1, X_2, \ldots, X_n$ be random variables with marginal distribution functions known as $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$, then*

$$X_1 + X_2 + \cdots + X_n \leq_{cx} F_{X_1}^{-1}(U) + F_{X_2}^{-1}(U) + \cdots + F_{X_n}^{-1}(U), \tag{2}$$

and

$$X_1 + X_2 + \cdots + X_n \geq_{cx} \mathrm{E}[X_1|\Lambda] + \mathrm{E}[X_2|\Lambda] + \cdots + \mathrm{E}[X_n|\Lambda], \tag{3}$$

with U a uniform(0,1) distributed random variable, and with Λ an arbitrary variable for which the conditional distributions of X_i given Λ are known.
The upper bound of (2) can be improved to a closer bound

$$X_1 + X_2 + \cdots + X_n \leq_{cx} F_{X_1|\Lambda}^{-1}(U) + F_{X_2|\Lambda}^{-1}(U) + \cdots + F_{X_n|\Lambda}^{-1}(U), \tag{4}$$

with U and Λ as before.

Note that the lower bound of (3) and the improved upper bound of (4) perform better the more Λ resembles the original sum.

If we define the inverse distribution as $F_{X_i}^{-1}(p) = \inf\{x \in \mathbb{R} : F_{X_i}(x) \geq p\}$, and $F_{X_i}^{-1+}(p) = \sup\{x \in \mathbb{R} : F_{X_i}(x) \leq p\}$, $p \in [0,1]$, the results of theorem 2.1 can be extended to functions of the variables X_i, by making use of the following lemma:

Lemma 2.2 *If ψ is a continuous real-valued function and p is any number in $]0,1[$, then if ψ is non-decreasing, $F_{\psi(X)}^{-1}(p) = \psi(F_X^{-1}(p))$, and if ψ is non-increasing, $F_{\psi(X)}^{-1}(p) = \psi(F_X^{-1+}(1-p))$.*

2.3. Stochastic interest rate model

As mentioned in the introduction, there exists a long list of stochastic processes, useful to model interest rates. In the sequel we will give an elaborated example of our method for a well known and frequently used easy model, the Brownian motion with drift, defined by the stochastic differential equation

$$dX(t) = \mu dt + \sigma dW(t),$$

with $W = \{W(t)\}$ a standard Brownian motion.
This model benefits from the fact that it is one of the most easiest models to describe a stochastic interest rate. An advantage of this model can be found in the appropriateness for situations with rather great variation; a disadvantage however is that for long periods, a very large value (both positive and negative) could be reached, which imposes the possibility of instability. However, by implementing a restriction as suggested in subsection 2.1, this disadvantage can be perfectly avoided.

If we use the notation $F(t,x)$ for the cumulative distribution function of the variable $X(t)$, it is well known that

$$\begin{array}{ll} F(t,x) = \Phi\left(\frac{x-\mu t}{\sigma\sqrt{t}}\right), & x \in \mathbb{R} \\ F^{-1}(t,p) = \mu t + \sigma\sqrt{t}\Phi^{-1}(p), & p \in [0,1], \end{array}$$

with $\Phi(x)$ the standard normal cumulative distribution function.

3. CONSTANT ANNUITIES

In this section, we first present our results without specifying the stochastic process used to model the interest rate. We provide expressions for stochastic bounds to general constant annuities. Afterwards, we show how for each of these bounds analytical results can be obtained in the case of a Brownian motion with drift.

3.1. General case

Consider a discrete annuity over the time-interval $[0,t]$, with linear truncate stochastic interest rate with floor f and ceiling c as defined in definition 2.2:

$$V(t) = \sum_{i=1}^{n} e^{-\tilde{S}_f^c(t_i, X(t_i))}, \tag{5}$$

where $X = \{X(t)\}$ is a stochastic process with $X(t_i)$ denoting the compounded rate of return for the period $[0,t_i]$.

Applying the methodology of convex bounds (see subsection 2.2), the following results can be obtained straightforwardly:

Theorem 3.1 *The annuity of equation (5) can be bounded in convex ordering sense as*

$$V_{low}(t) \leq_{cx} V(t) \leq_{cx} V_{imupp}(t) \leq_{cx} V_{upp}(t),$$

where the stochastic bounds are determined by

$$\left\{ \begin{array}{rcl} V_{upp}(t) & = & \sum_{i=1}^{n} e^{-\tilde{S}_f^c(t_i, F^{-1+}_{X(t_i)}(1-U))} \\ V_{low}(t) & = & \sum_{i=1}^{n} \mathrm{E}[e^{-\tilde{S}_f^c(t_i, X(t_i))}|\Lambda] \\ V_{imupp}(t) & = & \sum_{i=1}^{n} e^{-\tilde{S}_f^c(t_i, F^{-1+}_{X(t_i)|\Lambda}(1-U))}. \end{array} \right.$$

In these expressions, U is a uniform(0,1) distributed random variable, and Λ is an arbitrary variable such that the distribution of $X(t_i)|\Lambda$ is known.

By taking limits, the case of a continuous annuity $V(t) = \int_0^t e^{-\tilde{\mathcal{S}}_f^c[\tau, X(\tau)]}\, d\tau$ can be solved in a similar way.

Remark: Note that each of the previous results remain valid when the linear truncate interest rate $\tilde{\mathcal{S}}_f^c[t, X(t)]$ is replaced by an ordinary truncate rate $\mathcal{S}_f^c[X(t)]$.

3.2. The case of a discrete annuity with a brownian motion

Consider a discrete annuity certain as in equation (5).
Since $X(t_i)$ corresponds to the *cumulative* interest rate for the period $[0, t_i]$, it can be written as $X(t_i) = Y(t_1) + \cdots + Y(t_i)$, with $Y(t_k)$ the interest rate for the period $[t_{k-1}, t_k]$. In the Brownian model, we assume that the vector $Y = (Y(t_1), Y(t_2), \ldots, Y(t_n))$ consists of independent normally distributed variables.
Next, define Λ as a lineair combination of the variables $Y(t_k)$, or

$$\Lambda = \sum_{i=1}^{n} a_i Y(t_i), \qquad\qquad a_i \in \mathbb{R},$$

such that the distribution function of Λ kan be written as

$$F_\Lambda(\lambda) = \Phi\left(\frac{\lambda - \mu t \sum a_i}{\sqrt{\sigma^2 t \sum a_i^2}}\right).$$

Since Λ and each variable $X(t_i)$ $(i = 1, ..., n)$ are combinations of the components of Y, it follows that $X(t_i)|\Lambda$ is also normally distributed with mean and variance given by

$$\begin{cases} \bar{\mu}_i(\Lambda) = \mathrm{E}[X(t_i)] + corr[X(t_i), \Lambda]\frac{\sigma_{X(t_i)}}{\sigma_\Lambda}(\Lambda - \mathrm{E}[\Lambda]) \\ \bar{\sigma}_i^2 = \sigma_{X(t_i)}^2(1 - corr[X(t_i), \Lambda]^2). \end{cases}$$

The following results hold :

Theorem 3.2 *In the Brownian case, the discrete annuity certain as in equation (5) can be bounded by*

$$V_{low}(t) \leq_{cx} V(t) \leq_{cx} V_{imupp}(t) \leq_{cx} V_{upp}(t),$$

where

$$V_{upp}(t) = \sum_{i=1}^{n} e^{-\tilde{\mathcal{S}}_f^c[\left(\mu t_i + \sigma\sqrt{t_i}\Phi^{-1}(1-U)\right)]},$$

$$V_{low}(t) = \sum_{i=1}^{n}\left(e^{-f\cdot t_i}\Phi\left(\frac{f\cdot t_i - \bar{\mu}_i(\Lambda)}{\bar{\sigma}_i}\right) + e^{-c\cdot t_i}\Phi\left(\frac{\bar{\mu}_i(\Lambda) - c\cdot t_i}{\bar{\sigma}_i}\right)\right.$$
$$\left. + e^{-\bar{\mu}_i(\Lambda)+\frac{1}{2}\bar{\sigma}_i^2}\cdot\left(\Phi\left(\frac{c\cdot t_i - \bar{\mu}_i(\Lambda) + \bar{\sigma}_i^2}{\bar{\sigma}_i}\right) - \Phi\left(\frac{f\cdot t_i - \bar{\mu}_i(\Lambda) + \bar{\sigma}_i^2}{\bar{\sigma}_i}\right)\right)\right),$$

and

$$V_{imupp} = \sum_{i=1}^{n} G_i(U, \Lambda),$$

where the functions $G_i : [0,1] \times \mathbb{R} \to \mathbb{R}^+ : (p,\lambda) \mapsto G_i(p,\lambda)$ *are defined by*

$$G_i(p,\lambda) = \begin{cases} e^{-c \cdot t_i} & \textit{if } \; p \in [0, p_i^{(2)}(\lambda)[\\ e^{-f \cdot t_i} & \textit{if } \; p \in [1 - p_i^{(1)}(\lambda), 1] \\ e^{\bar{\sigma}_i \Phi^{-1}(p) - \bar{\mu}_i(\lambda)} & \textit{if } \; p \in [p_i^{(2)}(\lambda), 1 - p_i^{(1)}(\lambda)[\end{cases}$$

with $p_i^{(1)}(\lambda) = \Phi(\frac{f \cdot t_i - \bar{\mu}_i(\lambda)}{\bar{\sigma}_i})$ *and* $p_i^{(2)}(\lambda) = \Phi(\frac{-c \cdot t_i + \bar{\mu}_i(\lambda)}{\bar{\sigma}_i})$*.*

Proof. This follows after a few calculations when the methodology explained in subsection 2.2 is applied. □

Concerning the distributions of these bounds, the results are summarized in the following theorem, where the notation F_Z is used as notation for the cumulative distribution function of the variable Z, or $F_Z(x) = \mathrm{Prob}\,(Z \leq x)$.

Theorem 3.3 *The cumulative distribution functions of the convex bounds of theorem 3.2 can be calculated as follows:*

$$\begin{cases} F_{V_{upp}}(x) &= 1 - \Phi(\nu_x), \\ F_{V_{low}}(x) &= 1 - \Phi\left(\frac{\lambda_x - \mu t \sum a_i}{\sqrt{\sigma^2 t \sum a_i^2}}\right), \\ F_{V_{imupp}}(x) &= \int_{-\infty}^{+\infty} \kappa(\lambda, x) dF_{\Lambda}(\lambda), \end{cases}$$

with ν_x*,* λ_x *defined implicitly and* $\kappa(\lambda, x)$ *defined explicitly as*

$$\begin{cases} \sum_{i=1}^{n} e^{-\tilde{\mathcal{S}}_f^c[t_i, \mu t_i + \sigma\sqrt{t_i}\nu_x]} = x, \\ V_{low}(t)|_{\Lambda=\lambda_x} = x, \\ \kappa(\lambda, x) = \sup\{p \in [0,1] \mid \sum_{i=1}^{n} G_i(p, \Lambda = \lambda) \leq x\}. \end{cases}$$

Proof. In order to prove these statements, use can be made of the results mentioned in subsection 2.2. Note that the values reached by the functions $G_i : [0,1] \times \mathbb{R} \to \mathbb{R}^+$ in fact can be written as

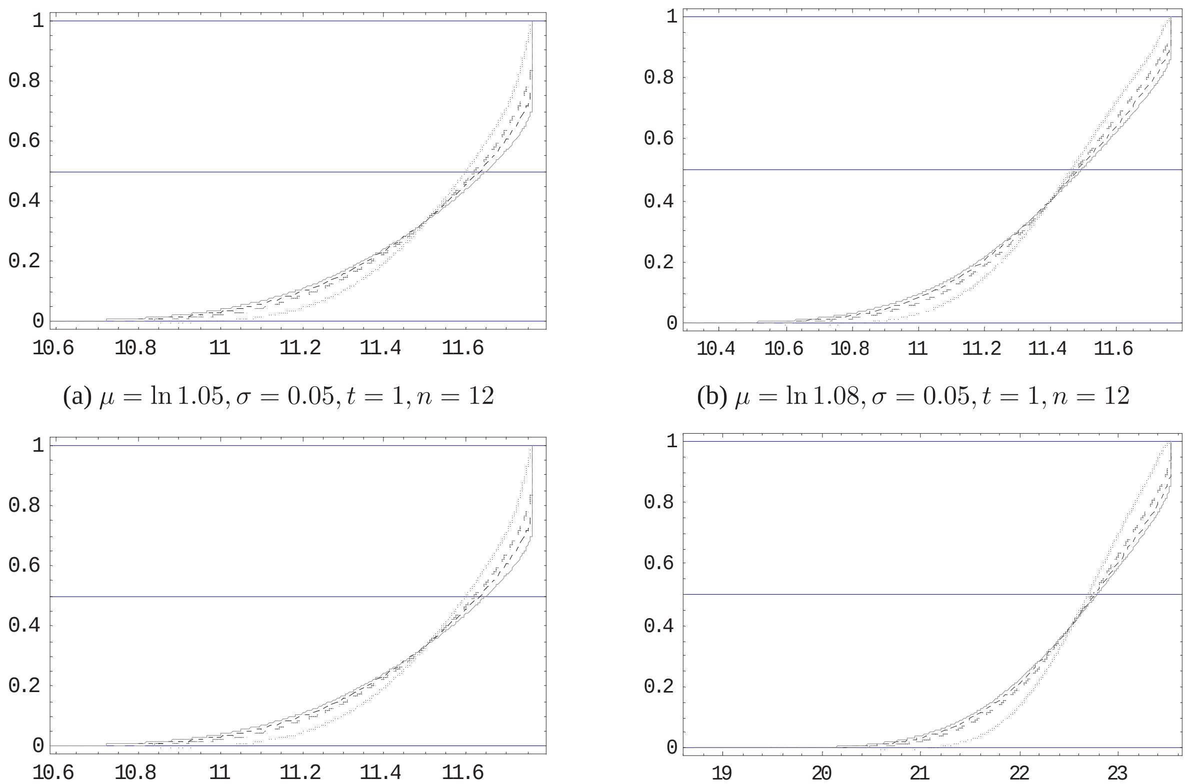
$$G_i(p,\lambda) = F^{-1}_{e^{-\tilde{\mathcal{S}}_f^c(t_i, X(t_i))}|\Lambda=\lambda}(p) = \mathrm{Prob}\left(e^{-\tilde{\mathcal{S}}_f^c(t_i, X(t_i))} \leq p \mid \Lambda = \lambda\right).$$

□

Remark: Note that — in analogy with the previous subsection — each of the previous results can be reformulated easily when the linear truncate interest rate $\tilde{\mathcal{S}}_f^c[t, X(t)]$ is replaced by an ordinary truncate rate $\mathcal{S}_f^c[X(t)]$.

Some numerical examples of these convex bounds are shown in figures 2 and 3, for different choices of the parameters. Both figures consist of four plots of the distribution function of the original discrete annuity of (5)(simulated by means of a Monte-Carlo procedure) and the distribution functions for the three convex bounds as obtained in theorem 3.3. Figure 2 deals with the case of an ordinary truncate stochastic interest rate, while in figure 3 the plots are made for linear truncate stochastic interest rates.

upper (__), lower (__), improved upper(- -), simulated (- -) bound



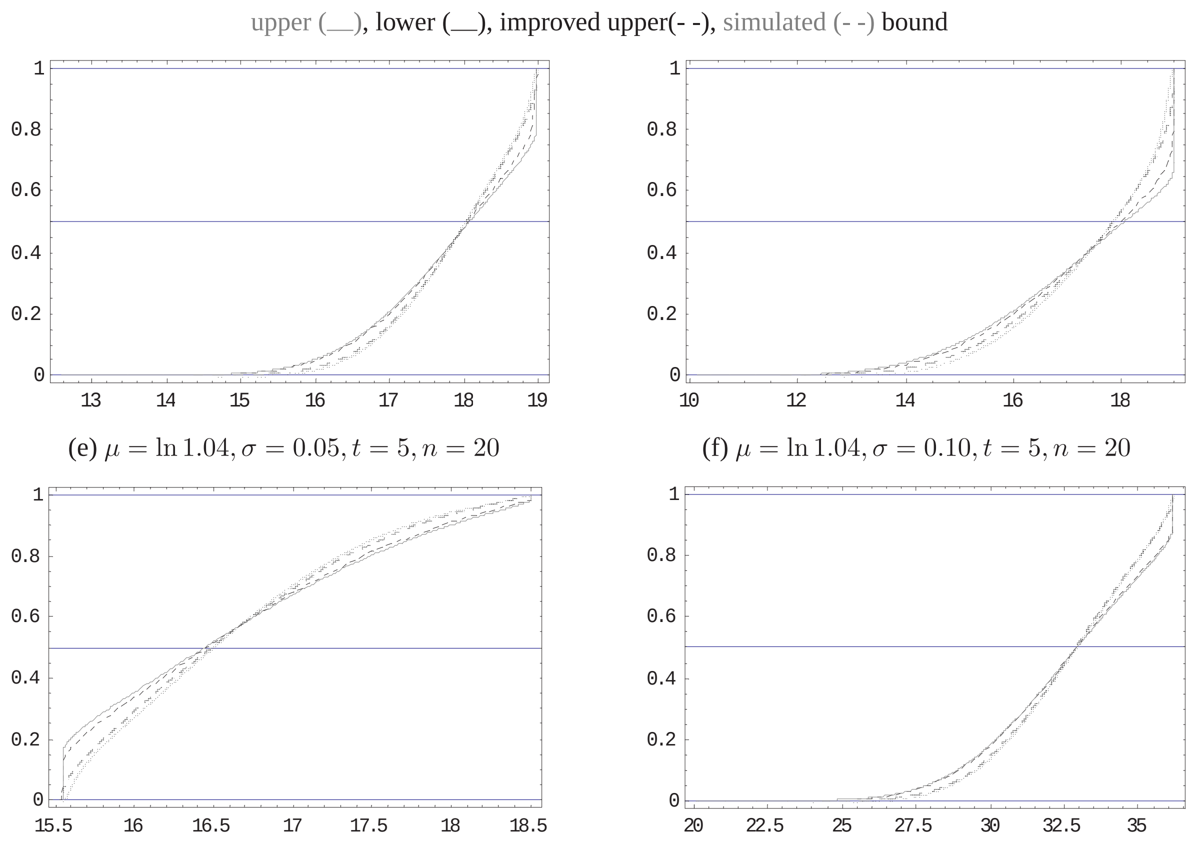
(a) $\mu = \ln 1.05, \sigma = 0.05, t = 1, n = 12$ (b) $\mu = \ln 1.08, \sigma = 0.05, t = 1, n = 12$

(c) $\mu = \ln 1.05, \sigma = 0.10, t = 1, n = 12$ (d) $\mu = \ln 1.05, \sigma = 0.05, t = 2, n = 24$

Figure 2: Examples of annuities with Brownian motion, truncate interest rate

The four plots in figure 2 are considered in a Brownian context, where we change in each plot the values for one of the parameters μ, σ, t and n. The conditioning variable Λ is defined by its coefficients $a_i = 1 + \frac{i}{24}$; for the floor and ceiling the parameter values are $f = 0.02$ and $c = 0.3$. It can be seen that the improved upper bound and the lower bound are close to the simulation of the distribution. The bounds are more accurate the lower the volatility σ. Note in each plot the kink in the distribution functions, the position of which is proportional to the probability that the stochastic interest rate is smaller than f (in the case of a kink on the right) or greater than c (in the case of a kink on the left).

Similar results about the performances of the bounds can be observed in the plots of figure 3, where we used a lineair truncate interest rate and a longer time horizon. The conditioning variable Λ here is defined by its coefficients $a_i = 20 - i/2$, with $i = 1, \ldots, 20$ for plots (e),(f) and (g) and $i = 1, \ldots, 40$ for plot (h). For the floor and ceiling, the values are $f = 0.02$ and $c = 0.3$ for the plots (e), (f), (h) and $f = 0.03$ and $c = 0.1$ for plot (g).



(e) $\mu = \ln 1.04, \sigma = 0.05, t = 5, n = 20$

(f) $\mu = \ln 1.04, \sigma = 0.10, t = 5, n = 20$

(g) $\mu = \ln 1.08, \sigma = 0.05, t = 5, n = 20$

(h) $\mu = \ln 1.04, \sigma = 0.05, t = 10, n = 40$

Figure 3: Examples of annuities with Brownian motion, lineair truncate interest rate

4. APPLICATIONS AND EXTENSIONS

Consider a general discrete annuity over the time-interval $[0, t]$, with a linear truncate stochastic interest rate with floor f and ceiling c as defined in definition 2.2:

$$V^*(t) = \sum_{i=1}^{n} \alpha(t_i)\, e^{-\tilde{S}_f^c(t_i, X(t_i))}, \tag{6}$$

where $X = \{X(t)\}$ is a stochastic process with $X(t_i)$ denoting the compounded rate of return for the period $[0, t_i]$. Theorem 3.1 can be extended as follows:

Theorem 4.1 *The annuity of equation (6) can be bounded in convex ordering sense as*

$$V^*_{low}(t) \leq_{cx} V^*(t) \leq_{cx} V^*_{imupp}(t) \leq_{cx} V^*_{upp}(t), \tag{7}$$

where the stochastic bounds are determined by

$$\begin{cases} V^*_{upp}(t) &= \sum_{i=1}^{n} \max(0,\alpha(t_i))\ e^{-\tilde{\mathcal{S}}^c_f(t_i,F^{-1+}_{X(t_i)}(1-U))} - \sum_{i=1}^{n} \min(0,-\alpha(t_i))\ e^{-\tilde{\mathcal{S}}^c_f(t_i,F^{-1}_{X(t_i)}(U))} \\ V^*_{low}(t) &= \sum_{i=1}^{n} \alpha(t_i)\ \mathrm{E}[e^{-\tilde{\mathcal{S}}^c_f(t_i,X(t_i))}|\Lambda] \\ V^*_{imupp}(t) &= \sum_{i=1}^{n} \max(0,\alpha(t_i))\ e^{-\tilde{\mathcal{S}}^c_f(t_i,F^{-1+}_{X(t_i)|\Lambda}(1-U))} - \sum_{i=1}^{n} \max(0,-\alpha(t_i))\ e^{-\tilde{\mathcal{S}}^c_f(t_i,F^{-1}_{X(t_i)|\Lambda}(U))}, \end{cases}$$

where U is a uniform(0,1) distributed random variable, and where Λ is an arbitrary variable such that the distribution of $X(t_i)|\Lambda$ is known.

Applications of this more general result are obvious, e.g.

- for an indexed payment, use can be made of $\alpha(t) = (1 + d_t)^t$, with d_t the indexing factor for the period $[t_{i-1}, t_i]$;
- for a life annuity, $\alpha(t) = {}_tp_x$, where ${}_tp_x$ is the classical notation used for the probability of a person of age x to be still alive after t years;
- for an indexed life annuity: $\alpha(t) = (1 + d_t)^t \cdot {}_tp_x$;
- for a life assurance policy: $\alpha(t) = {}_tp_x \cdot \mu_{x+t}$, where μ_x is the mortality intensity at age x.

In figure 4, we illustrate the possibilities of these applications. The four plots deal with the distribution function of the present value of

(i) an indexed payment, yearly 3%, 24 payments;

(j) a life annuity, age 35, 20 payments;

(k) an indexed life annuity, yearly 1.5%, age 30, 10 payments;

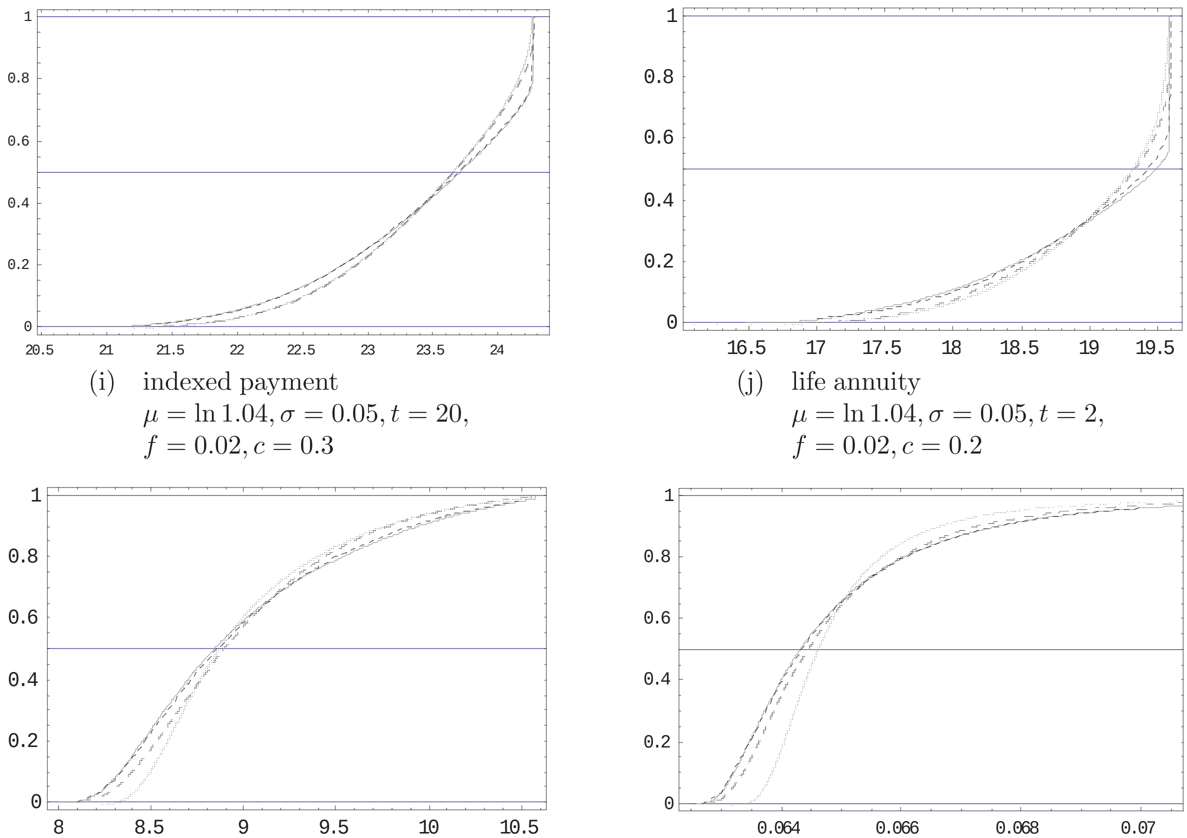
(l) a life assurance policy, age 40, duration of 20 years.

In order to conclude, we would like to mention that these results can be nicely extended, mainly in two directions. Firstly, the underlying stochastic process used to model the interest rates, can be modified. The use of a Vasicek or Ho-Lee model e.g. instead of a Brownian motion, seems to be more realistic. Secondly, also the function $\tilde{\mathcal{S}}^c_f$ can be altered, in order to deal with specific economic prerequisites, e.g. certain amortization schemes. Results about these and similar generalizations will be presented in forthcoming papers.

References

[1] De Schepper A., Goovaerts M.J. and Kaas R. (1997): A recursive scheme for perpetuities with random positive interest rates. *Scandinavian Actuarial Journal*, **1**, 1-10.

upper (__), lower (__), improved upper(- -), simulated (- -) bound



(i) indexed payment
$\mu = \ln 1.04, \sigma = 0.05, t = 20,$
$f = 0.02, c = 0.3$

(j) life annuity
$\mu = \ln 1.04, \sigma = 0.05, t = 2,$
$f = 0.02, c = 0.2$

(k) indexed life annuity
$\mu = \ln 1.04, \sigma = 0.1, t = 1,$
$f = 0.02, c = 0.2$

(l) life assurance policy
$\mu = \ln 1.04, \sigma = 0.05, t = 10,$
$f = 0.02, c = 0.3$

Figure 4: Examples of applications

[2] Dhaene J., Denuit M., Goovaerts M.J., Kaas R. and Vyncke D. (2002): The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics & Economics*, **31**(1), 3-34.

[3] Dhaene J., Denuit M., Goovaerts M.J., Kaas R. and Vyncke D. (2002): The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics & Economics*, **31**(2), 133-162.

[4] Goovaerts M.J., Dhaene J. and De Schepper A. (2000): Stochastic Upper Bounds for Present Value Functions. *Journal of Risk and Insurance*, **67**(1), 1-14.

FORWARD SWAP MARKET MODELS WITH JUMPS

Jan Liinev[†] **and Ernst Eberlein**[‡]

[†]*Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281, S9, 9000 Gent, Belgium*
[‡]*Department of Mathematical Stochastics, University of Freiburg, Freiburg, Germany*
Email: jan.liinev@ugent.be, eberlein@stochastik.uni-freiburg.de

Abstract

We consider the modelling of forward swap rates when the driving process is a general Lévy process. We present two ways of modelling these interest rates and show also how swaptions can be priced using bilateral Laplace transforms.

1. INTRODUCTION

The swaption market is one of the main interest rate markets. The models for forward swap rates in pure diffusion (Brownian motion) setting were developed by Jamshidian (1997) and Rutkowski (1999, 2001). More recent approaches for interest rate models involve jump-diffusions and more generally, models driven by Lévy processes. The latter are becoming increasingly popular in finance since they allow for greater flexibility compared to classical diffusion models (see e.g. Eberlein (2001)). A Lévy process $L = (L_t)_{t\geq 0}$ is a continuous in probability, càdlàg[1] stochastic process with independent and stationary increments. We denote the left-hand limit at t by $L_{t-} := \lim_{s\uparrow t} L_s$. The jump of a process at t is defined as $\Delta L_t = L_t - L_{t-}$. The distribution of a Lévy process is uniquely determined by any of its one-dimensional marginal distributions P^{L_t}, say by P^{L_1}, which is infinitely divisible.

In the context of instantaneous, continuously compounded interest rates, Björk, Kabanov and Runggaldier (1997) extend the classical Heath, Jarrow and Morton (1992) (henceforth HJM) framework to the case of a diffusion-multivariate point process, and Björk, Di Masi, Kabanov and Runggaldier (1997) to general semimartingales. Glasserman and Kou (2003) characterized the arbitrage-free dynamics of interest rates when the term structure is modelled through forward Libor rates or forward swap rates, in presence of both jumps and diffusion. They consider the case when a jump process is modelled through a finite number of marked point processes, in which

[1]meaning: right-continuous sample paths with existing left-hand limits

case the purely discontinuous part is of bounded variation. More explicitly, they place themselves into the generalized HJM framework of Björk, Kabanov and Runggaldier (1997) and show that the simple forward rates can be embedded in an arbitrage-free model of instantaneous forward rates. Eberlein and Özkan (2002) push the approach of Glasserman and Kou (2003) for forward Libor rates further into a more general setting of jump measures. Apart from the HJM framework of Björk, Di Masi, Kabanov and Runggaldier (1997), they consider the Lévy setting of Eberlein and Raible (1999). They show that the Lévy term structure approach to Libor markets can be embedded in the very general semimartingale approach of Jamshidian (1999). In addition, they construct the discrete tenor Lévy Libor model directly through backward induction, whence extending the approach of Musiela and Rutkowski (1997a, 1997b) from the case of pure diffusion to this Lévy setting.

In turn, we will develop a model of the forward swap rates by allowing the driving process to be a Lévy process. In that sense we slightly generalize the corresponding result in Glasserman and Kou (2003). However, our approach differs from that of Glasserman and Kou (2003) in a way that we do not start by showing that the forward swap rate model can be embedded in the framework of instantaneous forward rates of Björk, Di Masi, Kabanov and Runggaldier (1997) or Eberlein and Raible (1999). In fact, as pointed out in Hunt and Kennedy (2000), the extra burden of proving that the models fall within the HJM class is unnecessary. Instead, we use a numéraire-based approach and hence we do not explicitly specify the dynamics for the instantaneous forward rates or the bond prices. The outline of such a modelling approach in a pure diffusion setting can be found for instance in Pelsser (2000), and in Hunt and Kennedy (2000). Furthermore, we extend the backward induction method of Rutkowski (1999, 2001) to the case when the forward swap rates are driven by a general Lévy process.

We assume in the sequel that we are given a complete probability space $(\Omega, \mathcal{F}, P)$ equipped with filtration $(\mathcal{F}_t)_{t\geq 0}$, such that the filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\geq 0})$ satisfies the usual conditions, cf. Jacod and Shiryaev (1987).

In what follows, we consider a family of forward swap rates $S_i(t) := S(t, T_i, T_N)$ which have the same maturity date T_N for all $i = 0, \dots, N-1$

$$S_i(t) = \frac{B(t,T_i) - B(t,T_N)}{C_{i,N}(t)}, \tag{1}$$

where $B(t,T)$ denotes the time t price of a zero-coupon bond maturing at T. The accrual factor for any individual swap rate from such a family is given by

$$C_{i,N}(t) := \sum_{j=i+1}^{N} \delta_j B(t, T_j).$$

2. SWAP MARKET MODELS BASED ON A SINGLE MEASURE

Our aim is to develop an arbitrage-free model for the term structure of interest rates specified through forward swap rates $S_i(\cdot)$ under a single measure, namely T_N-forward measure P_N (also called terminal measure [2]). This is a useful approach if we want to determine the price of more complicated derivatives such as barrier swaptions where the pricing is done with respect to the collection of swap rates which reset on different dates but have a common maturity date. We extend the approach found in Hunt and Kennedy (2000), and in Pelsser (2000). We start by specifying the dynamics for the forward swap rates and then determine the necessary relationship for any corresponding term structure model to be arbitrage-free. Here we do not explicitly assume that the driving process is a Lévy process. However, the latter can be embedded in the given setting.

We assume that the tenor structure $0 < T_0 < T_1 < \ldots < T_N$ is given, and $\delta_j = T_j - T_{j-1}$ for $j = 1, \ldots, N$. By choosing the bond with the largest maturity T_N to be a numéraire, the discounted accrual factor (w.r.t. that numéraire) reads

$$P_t^i := \sum_{j=i+1}^{N} \frac{\delta_j B(t,T_j)}{B(t,T_N)} = \frac{C_{i,N}(t)}{B(t,T_N)}. \tag{2}$$

We also define for $0 \le i \le N-1$ the following product

$$\Psi_t^i := \prod_{j=0}^{i} \left(1 + \delta_{j+1} S_{j+1}(t)\right). \tag{3}$$

We follow the convention that empty sums and products denote zero and one, respectively. Note that $P_t^N \equiv S_N \equiv 0$. By using (1), we can express (2) through the recursive relation

$$P_t^i = \delta_{i+1} + (1 + \delta_{i+1} S_{i+1}(t)) \cdot P_t^{i+1}, \tag{4}$$

for $i = 0, \ldots, N-1$. Multiplying both sides of the equation (4) by Ψ_t^{i-1}, we obtain by backward induction, down from $i = N-1$, the non-recursive expression for P_t^i:

$$P_t^i = \frac{\sum_{j=i}^{N-1} \delta_{j+1} \Psi_t^{j-1}}{\Psi_t^{i-1}} = \sum_{j=i}^{N-1} \delta_{j+1} \prod_{k=i+1}^{j} (1 + \delta_k S_k(t)). \tag{5}$$

The next theorem states the forward swap rate model under the terminal measure P_N, and slightly generalizes Theorem 5.1 in Glasserman and Kou (2003).

Theorem 2.1 *For each $i = 0, \ldots, N-1$, let $\theta_i(\cdot)$ be a bounded $\mathbb{R}^d$-valued function and $G_i : \mathbb{R}_+ \times \mathbb{R}^r \to (-1, \infty)$ be a deterministic function in $G_{\mathrm{loc}}(\mu)$ [3]. Let W^N be a standard Brownian motion in $\mathbb{R}^d$ with respect to P_N, and μ the jump measure of a semimartingale with the continuous compensator $\nu^N(dt,dx) = \lambda^N(t,dx)dt$. The dynamics of $S_i(\cdot)$, $i = 0, \ldots, N-1$, is assumed to satisfy*

$$\frac{dS_i(t)}{S_i(t-)} = \alpha_i(t)dt + \theta_i(t)dW_t^N + \int_{\mathbb{R}^r} G_i(t,x)(\mu - \nu^N)(dt,dx). \tag{6}$$

Then this model is arbitrage-free if

$$\begin{aligned} \alpha_i(t) = &- \sum_{j=i+1}^{N-1} \frac{\delta_j \sum_{k=j}^{N-1} \delta_{k+1} \prod_{l=i+1}^{k} (1 + \delta_l S_l(t-)) S_j(t-) \theta_j(t) \theta_i(t)}{\sum_{k=i}^{N-1} \delta_{k+1} \prod_{l=i+1}^{k} (1 + \delta_l S_l(t-)) \cdot (1 + \delta_j S_j(t-))} \\ &+ \int_{\mathbb{R}^r} G_i(t,x) \left[1 - \frac{\sum_{j=i}^{N-1} \delta_{j+1} \prod_{k=i+1}^{j} \left(1 + \delta_k S_k(t-)(1 + G_k(t,x))\right)}{\sum_{j=i}^{N-1} \delta_{j+1} \prod_{k=i+1}^{j} (1 + \delta_k S_k(t-))} \right] \lambda^N(t,dx). \end{aligned} \tag{7}$$

[2] This is the measure associated to the numéraire bond price $B(t, T_N)$.
[3] For the definition of this set we refer to Jacod and Shiryaev (1987) II.1.27

Proof. As we want the model (6) to be arbitrage-free, each of the P_t^i, $i = 0, \ldots, N-1$, defined in (2), has to be a local martingale under the measure P_N. This imposes a relationship between the finite variation terms and the diffusion coefficients in (6), which we now derive.
Applying Itô's product rule to (4) yields

$$dP_t^i = (1 + \delta_{i+1} S_{i+1}(t-))\, dP_t^{i+1} + \delta_{i+1} P_{t-}^{i+1} dS_{i+1}(t) + \delta_{i+1} d\left[S_{i+1}, P^{i+1}\right]_t. \tag{8}$$

The quadratic covariation term on the right-hand side of (8) can be written as

$$\left[S_{i+1}, P^{i+1}\right]_t = \langle S_{i+1}^c, P^{i+1,c}\rangle_t + \sum_{0 \le s \le t} \Delta S_{i+1}(s) \Delta P_s^{i+1},$$

where $\Delta S_{i+1}(t)$ and ΔP_t^{i+1} denote the jumps of $S_{i+1}(t)$ and P_t^{i+1}, respectively. The superscript c indicates that we consider the process with continuous sample paths.
Recall that P_t^i has to be a local martingale under the measure P_N. Equating the local martingale parts of the SDE (8) while invoking (6) yields

$$\begin{aligned} dP_t^i =& (1 + \delta_{i+1} S_{i+1}(t-))\, dP_t^{i+1} + P_{t-}^{i+1} \delta_{i+1} S_{i+1}(t-) \theta_{i+1}(t) dW_t^N + \delta_{i+1} S_{i+1}(t-) \times \\ & \times \int_{\mathbb{R}^r} G_{i+1}(t,x) \left[\sum_{j=i+1}^{N-1} \delta_{j+1} \prod_{k=i+2}^{j} \big(1 + \delta_k S_k(t-)(1 + G_k(t,x))\big) \right] (\mu - \nu^N)(dt,dx) \end{aligned} \tag{9}$$

In order to obtain a non-recursive expression for dP_t^i, we multiply both sides of the equation (9) by Ψ_{t-}^{i-1}, and proceed by backward induction, down from $i = N-1$. It can then be shown that the diffusion term in that non-recursive SDE equals

$$P_{t-}^i \sum_{j=i+1}^{N-1} \frac{\Psi_{t-}^{j-1} P_{t-}^j}{\Psi_{t-}^{i-1} P_{t-}^i} \left(\frac{\delta_j S_j(t-)}{1 + \delta_j S_j(t-)} \right) \theta_j(t) dW_t^N. \tag{10}$$

Equating the finite variation terms in (8) yields

$$\begin{aligned} & \delta_{i+1} P_{t-}^{i+1} \alpha_{i+1}(t) S_{i+1}(t-) dt + \delta_{i+1} d\langle S_{i+1}^c, P^{i+1,c}\rangle_t + \delta_{i+1} S_{i+1}(t-) \times \\ & \times \int_{\mathbb{R}^r} G_{i+1}(t,x) \left[\sum_{j=i+1}^{N-1} \delta_{j+1} \prod_{k=i+2}^{j} \big(1 + \delta_k S_k(t-)(1 + G_k(t,x))\big) - P_{t-}^{i+1} \right] \nu^N(dt,dx) = 0, \end{aligned} \tag{11}$$

where from (6) and (10)

$$d\langle S_{i+1}^c, P^{i+1,c}\rangle_t = S_{i+1}(t-) \theta_{i+1}(t) P_{t-}^{i+1} \sum_{j=i+2}^{N-1} \frac{\Psi_{t-}^{j-1} P_{t-}^j}{\Psi_{t-}^{i-1} P_{t-}^{i+1}} \left(\frac{\delta_j S_j(t-)}{1 + \delta_j S_j(t-)} \right) \theta_j(t) dt. \tag{12}$$

Combining (11) and (12) by taking into account the definition (3) and relation (5), we can easily express the drift term α_i in (6) through forward swap rates and their volatilities, yielding (7). □

Similarly, one can construct another type of forward swap rate model for so-called reverse swap markets where the family of swap rates to be modelled has a common start date and different maturities. Such type of swap rates underlie for example the spread options. In pure diffusion setting this model is dealt with in Hunt and Kennedy (2000), and Pelsser (2000). It is also possible to extend this model into semimartingale setting, see Liinev (2003) for details.

3. THE DISCRETE TENOR LÉVY SWAP RATE MODEL

We make the following assumptions concerning the dynamics of the forward swap rates. Let μ^L be the jump measure of a Lévy process L. We also assume that the initial term structure of interest rates, specified by bond prices $B(0,T_j)$, $j=0,\dots,N$, is given and that $B(0,T_j)$ are strictly decreasing in the second variable, i.e. $B(0,T_j) > B(0,T_{j+1})$, $j=0,\dots,N-1$.

Assumption 3.1
For any maturity T_i, $i=0,\dots,N-1$, there exists a function $\gamma_1(\cdot,\cdot,T_i):\Omega\times[0,T_i]\to\mathbb{R}_+$, and a function $\gamma_2(\cdot,\cdot,\cdot,T_i):\Omega\times\mathbb{R}\times[0,T_i]\to\mathbb{R}_+$, both predictable with $\gamma_2\in G_{\mathrm{loc}}(\mu^L)$, such that

$$dS(t,T_i,T_N)=S(t-,T_i,T_N)\left(\gamma_1(t,T_i)dW_t^{T_{i+1}}+\int_{\mathbb{R}}\gamma_2(x,t,T_i)(\mu^L-\nu^{T_{i+1},L})(dt,dx)\right), \tag{13}$$

where $W^{T_{i+1}}$ is a $P_{T_{i+1}}$-standard Brownian motion and $\nu^{T_{i+1},L}(dt,dx)=\nu^{T_{i+1}}(dx)dt$ is the $P_{T_{i+1}}$-compensator of μ^L. We assume that $\nu^{T_{i+1},L}$ satisfies the integrability condition $\int_{|x|>1}\exp(ux)\nu^{T_{i+1}}(dx)<\infty$, for $-M\le u\le M$, where M is a positive constant. To guarantee that the swap rate is positive we assume that $\gamma_2(\Delta L_t,t,T_i)\mathbb{1}_{\Delta L_t\neq 0}>-1$.

We also assume that the functions γ_2 and γ_1 satisfy the integrability conditions $\int_0^{T_i}\int_{\mathbb{R}}\left(\sqrt{\gamma_2(x,t,T_i)+1}-1\right)^2\nu^{T_{i+1},L}(dt,dx)<\infty$ a.s. and $\int_0^{T_i}(\gamma_1(t,T_i))^2(dt)<\infty$ a.s., respectively, and that the initial condition for (13) is given by $S(0,T_i,T_N)=\frac{B(0,T_i)-B(0,T_N)}{C_{i,N}(0)}$.

In the following section we show how to construct measures $P_{T_{i+1}}$ such that (13) in Assumption 3.1 is satisfied.

3.1. Construction of the forward swap measures

We follow Rutkowski (1999, 2001) in order to construct an arbitrage-free bond market which is based on the Lévy swap rate model. We consider again the family of forward swap rates $S(t,T_i):=S(t,T_i,T_N)$ for $i=0,\dots,N-1$ which have a common expiration date T_N, but differ in length of the underlying swap agreement. The essence of this approach is that by fixing T_N, one starts the construction of the model backwards in terms of maturities (thus, starting from the largest maturity), specifying at each step the change of measure under which the swap rate in the following step is a local martingale.

We assume that the tenor structure $0<T_0<T_1<\dots<T_N=T^*$ is given, and $\delta_i=T_i-T_{i-1}$ for $i=1,\dots,N$, δ_0 is the length of accrual period from settlement to T_0. Note that $T_i=\sum_{j=0}^{i}\delta_j$. For our construction we set $T_l^*=T_{N-l}$ and, in particular, $T^*:=T_0^*=T_N$. Thus, we consider a "reversed" tenor structure $0<T_N^*<T_{N-1}^*<\dots<T_1^*<T_0^*=T_N$.

Suppose for the moment that we are given a family of bond prices $B(t,T_m)$, $m=1,\dots,N$. We postulate that $P^*:=P_{T^*}$ is the forward measure for the date T^*, the process W^{T^*} is the corresponding Brownian motion, and $\nu^{T^*,L}$ the corresponding compensator of μ^L. For any $m=1,\dots,N-1$ the accrual factor is then given by

$$C_{N-m,N}(t)=\sum_{l=N-m+1}^{N}\delta_l B(t,T_l)=\sum_{k=0}^{m-1}\delta_{N-k}B(t,T_k^*),\ \forall t\in[0,T_{N-m+1}]. \tag{14}$$

Through $C_{N-m,N}(\cdot)$ we introduce the forward swap measure as follows. For a fixed $i = 0, \ldots, N$, a probability measure $\widetilde{P}_{T_i}$, equivalent to P^*, is called the fixed-maturity forward swap measure for the date T_i if for every $k = 0, \ldots, N$, the relative bond price

$$Z_{N-i+1}(t, T_k) := \frac{B(t, T_k)}{C_{i-1,N}(t)} = \frac{B(t, T_k)}{\delta_i B(t, T_i) + \cdots + \delta_N B(t, T_N)} \tag{15}$$

follows a local martingale under $\widetilde{P}_{T_i}$. Thus, the forward swap measure corresponds to the choice of the accrual factor as a numéraire asset. Put another way, for any fixed $m \in \{1, \ldots, N\}$, the relative bond prices

$$Z_m(t, T_k^*) = \frac{B(t, T_k^*)}{C_{N-m,N}(t)} = \frac{B(t, T_k^*)}{\delta_{N-m+1} B(t, T_{m-1}^*) + \ldots + \delta_N B(t, T^*)}, \quad t \in [0, T_k^* \wedge T_{m-1}^*], \tag{16}$$

follow local martingales under $\widetilde{P}_{T_{m-1}^*}$. For all $t \in [0, T_m^*]$ the forward swap rate for date T_m^* equals

$$S(t, T_m^*) = \frac{B(t, T_m^*) - B(t, T^*)}{\delta_{N-m+1} B(t, T_{m-1}^*) + \ldots + \delta_N B(t, T^*)} = Z_m(t, T_m^*) - Z_m(t, T^*), \tag{17}$$

for all $t \in [0, T_m^*]$. Therefore $S(\cdot, T_m^*)$ also follows a local martingale under $\widetilde{P}_{T_{m-1}^*}$.

Remark 3.1 *The relative bond price*

$$Z_1(t, T_k^*) = \frac{B(t, T_k^*)}{C_{N-1,N}(t)} = \frac{B(t, T_k^*)}{\delta_N B(t, T^*)} = \frac{1}{\delta_N} F_B(t, T_k^*, T^*), \tag{18}$$

where $F_B(t, T_k^*, T^*)$ stands for forward price, and thus the probability measure $\widetilde{P}_{T^*}$ coincides with the forward martingale measure P_{T^*}.

We proceed with the backward construction of forward swap measures. The first step is to introduce the forward swap rate for the date T_1^* by postulating (according to Assumption 3.1) that the forward swap rate $S(\cdot, T_1^*)$ solves the SDE

$$dS(t, T_1^*) = S(t, T_1^*) \left(\gamma_1(t, T_1^*) d\widetilde{W}_t^{T^*} + \int_{\mathbb{R}} \gamma_2(x, t, T_1^*)(\mu^L - \widetilde{\nu}^{T^*,L})(dt, dx) \right), \tag{19}$$

for all $t \in [0, T_1^*]$, where $\widetilde{W}^{T^*} = W^{T^*}$ and $\widetilde{\nu}^{T^*,L} = \nu^{T^*,L}$, with initial condition $S(0, T_1^*) = \frac{B(0,T_1^*) - B(0,T^*)}{\delta_N B(0,T^*)}$.

To specify the process $S(\cdot, T_2^*)$, we need first to introduce a forward swap measure $\widetilde{P}_{T_1^*}$. Referring to Remark 3.1 we have that $\widetilde{P}_{T^*} = P_{T^*}$, and $Z_1(\cdot, T_k^*)$ follows a (strictly) positive local martingale under $\widetilde{P}_{T^*}$,

$$\mathrm{d}Z_1(t, T_k^*) = Z_1(t-, T_k^*) \left(\xi_1(t, T_k^*) \, \mathrm{d}\widetilde{W}_t^{T^*} + \int_{\mathbb{R}} \xi_2(x, t, T_k^*) \left(\mu^L - \widetilde{\nu}^{T^*,L} \right) (dt, dx) \right). \tag{20}$$

By suitably rewriting (15) we can express the relative bond price $Z_2(\cdot, T_k^*)$ as

$$Z_2(t, T_k^*) = \frac{Z_1(t, T_k^*)}{\delta_{N-1} Z_1(t, T_1^*) + 1}.$$

According to the definition of a forward swap measure, we postulate that for every k

$$Z_2(t, T_k^*) = \frac{Z_1(t, T_k^*)}{\delta_{N-1} Z_1(t, T_1^*) + 1} \tag{21}$$

follows a local martingale under $\widetilde{P}_{T_1^*}$.

In order to find the dynamics of Z_2 under $\widetilde{P}_{T_1^*}$ we use the following lemma.

Lemma 3.1 *Let G, H be real-valued adapted processes under some probability measure P, satisfying the following SDEs*

$$\begin{aligned} dG_t &= g_1(t)\, dW_t + \int_{\mathbb{R}} g_2(t, x)\, (\mu^L - \nu^L)(dt, dx) \\ dH_t &= h_1(t)\, dW_t + \int_{\mathbb{R}} h_2(t, x)\, (\mu^L - \nu^L)(dt, dx) \end{aligned}$$

where W_t is P-Brownian motion and $\nu^L(dt, dx)$ is P-compensator of μ^L. Let g_1, h_1 be square-integrable P-a.s. and $g_2, h_2 \in G_{\mathrm{loc}}(\mu^L)$. Suppose $H_t > -1$. Define

$$Y_t := \frac{1}{1 + H_t} > 0.$$

Then the process YG has the local martingale dynamics

$$\begin{aligned} d(YG)_t &= Y_{t-}(g_1(t) - Y_{t-} G_{t-} h_1(t)) d\widetilde{W}_t \\ &\quad + \int_{\mathbb{R}} \left(\frac{G_{t-} + g_2(t, x)}{1 + H_{t-} + h_2(t, x)} - \frac{G_{t-}}{1 + H_{t-}} \right) (\mu^L - \widetilde{\nu}^L)(dt, dx) \end{aligned}$$

under a new measure $\widetilde{P}$, $\widetilde{P} \overset{\mathrm{loc}}{\ll} P$, and where $\widetilde{W}_t$ is $\widetilde{P}$-Brownian motion,

$$d\widetilde{W}_t = dW_t - Y_{t-} h_1(t)\, dt,$$

and $\widetilde{\nu}^L(dt, dx)$ is $\widetilde{P}$-compensator of μ^L given by

$$\widetilde{\nu}^L(dt, dx) = \frac{1 + H_{t-} + h_2(t, x)}{1 + H_{t-}} \nu^L(dt, dx).$$

Applying Lemma 3.1 to processes $G = Z_1(\cdot, T_k^*)$ and $H = \delta_{N-1} Z_1(\cdot, T_1^*)$, it is easy to see that for $Z_2(\cdot, T_k^*)$ to follow a local martingale under $\widetilde{P}_{T_1^*}$ it suffices to assume that the process $\widetilde{W}^{T_1^*}$ follows a Brownian motion under $\widetilde{P}_{T_1^*}$, and that $\widetilde{\nu}^{T_1^*, L}$ is a $\widetilde{P}_{T_1^*}$-compensator of μ^L. Note that from (17) and (18)

$$Z_1(t, T_1^*) = \frac{B(t, T_1^*)}{\delta_N B(t, T^*)} = S(t, T_1^*) + Z_1(t, T^*) = S(t, T_1^*) + \delta_N^{-1}. \tag{22}$$

Differentiating both sides of the last equality and invoking (19) and (20), we obtain

$$
\begin{aligned}
&Z_1(t-,T_1^*)\xi_1(t,T_1^*)\,\mathrm{d}\widetilde{W}_t^{T^*} + Z_1(t-,T_1^*)\int_{\mathbb{R}} \xi_2(x,t,T_1^*)\left(\mu^L - \widetilde{\nu}^{T^*,L}\right)(dt,dx) \\
&= S(t-,T_1^*)\gamma_1(t,T_1^*)d\widetilde{W}_t^{T^*} + S(t-,T_1^*)\int_{\mathbb{R}} \gamma_2(x,t,T_1^*)(\mu^L - \widetilde{\nu}^{T^*,L})(dt,dx).
\end{aligned}
$$

As the Gaussian and the jump part of a semimartingale do not interact (see Jacod and Shiryaev (1987) II.2.34), in order for this equality to hold we set

$$Z_1(t-,T_1^*)\xi_1(t,T_1^*) = S(t-,T_1^*)\gamma_1(t,T_1^*)$$

$$Z_1(t-,T_1^*)\xi_2(x,t,T_1^*) = S(t-,T_1^*)\gamma_2(x,t,T_1^*).$$

Consequently, $\widetilde{W}^{T_1^*}$ is explicitly given by the formula

$$\widetilde{W}_t^{T_1^*} = \widetilde{W}_t^{T^*} - \int_0^t \frac{\delta_{N-1}S(s,T_1^*)}{1+\delta_{N-1}\delta_N^{-1}+\delta_{N-1}S(s,T_1^*)}\cdot\gamma_1(s,T_1^*)ds,$$

and the $\widetilde{P}_{T_1^*}$-compensator of μ^L by

$$\widetilde{\nu}^{T_1^*,L} = \frac{1+\delta_{N-1}\delta_N^{-1}+\delta_{N-1}S(s,T_1^*)+S(s,T_1^*)\gamma_2(x,t,T_1^*)}{1+\delta_{N-1}\delta_N^{-1}+\delta_{N-1}S(s,T_1^*)}\widetilde{\nu}^{T^*,L}.$$

Now we can define, using Girsanov's theorem, the associated forward swap measure $\widetilde{P}_{T_1^*}$ (through Lemma 3.1).

Subsequently, we introduce the process $S(t,T_2^*)$ by postulating that it solves the SDE

$$dS(t,T_2^*) = S(t-,T_2^*)\left(\gamma_1(t,T_2^*)d\widetilde{W}_t^{T_1^*} + \int_{\mathbb{R}}\gamma_2(x,t,T_2^*)(\mu^L - \widetilde{\nu}^{T_1^*,L})(dt,dx)\right),$$

for all $t\in[0,T_2^*]$ with the initial condition $S(0,T_2^*) = \frac{B(0,T_2^*)-B(0,T^*)}{\delta_{N-1}B(0,T_1^*)+\delta_N B(0,T^*)}$.

In the next inductive step we are looking for $S(t,T_3^*)$ by considering the process $Z_3(t,T_k^*)$, and consequently define $\widetilde{P}_{T_2^*}$. Extension to the general case, where we would like to determine the forward swap measure $\widetilde{P}_{T_m^*}$, and the forward swap rate $S(\cdot,T_{m+1}^*)$ is straightforward, see Liinev (2003) for details.

3.2. Special cases

We now turn to the special case where γ_1 and γ_2 in (13) are deterministic. In this case we can model the swap rates directly through the driving Lévy process.

Assume that there exists a constant $c\geq 0$ and a function γ on $[0,T]$ such that

$$\gamma_1(t,T_i) = \sqrt{c}\gamma(t,T_i),\ \gamma_2(x,t,T_i) = \gamma(t,T_i)x.$$

Then $\widetilde{L}_t^{T_{i+1}} := \sqrt{c}\widetilde{W}_t^{T_{i+1}} + \int_0^t \int_{\mathbb{R}} x \left(\mu^L - \widetilde{\nu}^{T_{i+1},L}\right)(ds, dx)$ is a Lévy process under $\widetilde{P}_{T_{i+1}}$. The dynamics of the forward swap rate is driven by the Lévy process $\widetilde{L}^{T_{i+1}}$:

$$dS(t, T_i, T_N) = S(t-, T_i, T_N)\gamma(t, T_i)d\widetilde{L}_t^{T_{i+1}}.$$

We can write this as a stochastic exponential

$$S(t, T_i, T_N) = S(0, T_i, T_N)\mathcal{E}\left(\int_0^{\cdot} \gamma(s, T_i)d\widetilde{L}_s^{T_{i+1}}\right)_t.$$

In order to ensure that the swap rates are positive we have to assume that the jumps of $\int_0^{\cdot} \gamma(s, T_i)d\widetilde{L}_s^{T_{i+1}}$ are strictly larger than -1. However, this can be replaced by the following alternative assumption:

$$S(t, T_i, T_N) = S(0, T_i, T_N)\exp\left(\int_0^t \gamma(s, T_i)d\widetilde{L}_s^{T_{i+1}}\right), \tag{23}$$

where γ is a positive deterministic function such that $\int_0^t (\gamma(s, T_i))^2 \, ds < \infty$. We also assume that there exists $c \geq 0$, and a continuously differentiable function $b : \mathbb{R}_+ \to \mathbb{R}$ such that

$$\widetilde{L}_t^{T_{i+1}} - b(t) = \sqrt{c}\widetilde{W}_t^{T_{i+1}} + \int_0^t \int_{\mathbb{R}} x \left(\mu^L - \widetilde{\nu}^{T_{i+1},L}\right)(ds, dx) \tag{24}$$

is a Lévy process under $\widetilde{P}_{T_{i+1}}$ and $\widetilde{\nu}^{T_{i+1},L}$ denotes the Lévy measure of $\widetilde{L}_1^{T_{i+1}} - b(1)$, and that

$$\begin{aligned}\int_0^t \gamma(s, T_i)b'(s)ds = -\Bigg(&\int_0^t \int_{\mathbb{R}} \left(\mathrm{e}^{\gamma(s,T_i)x} - 1 - \gamma(s, T_i)x\right)\widetilde{\nu}^{T_{i+1},L}(ds, dx) \\ &+ \frac{c}{2}\int_0^t (\gamma(s, T_i))^2 \, ds\Bigg).\end{aligned} \tag{25}$$

According to Özkan (2002) Lemma 4.7, the assumptions (23)-(24) and the condition (25) are necessary in order $S(\cdot, T_i, T_N)$ to be a martingale under $\widetilde{P}_{T_{i+1}}$. It can also be shown that the modelling approach (23) is equivalent to

$$\begin{aligned}S(t, T_i, T_N) = S(0, T_i, T_N)\mathcal{E}\Bigg(&\sqrt{c}\int_0^{\cdot} \gamma(s, T_i)d\widetilde{W}_s^{T_{i+1}} \\ &+ \int_0^{\cdot} \int_{\mathbb{R}} \left(\mathrm{e}^{\gamma(s,T_i)x} - 1\right)\left(\mu^L - \widetilde{\nu}^{T_{i+1},L}\right)(ds, dx)\Bigg)_t.\end{aligned}$$

4. NOTE ON PRICING OF SWAPTIONS

By using general valuation results (see e.g. Musiela and Rutkowski (1997b)), the time $t = 0$ price of the forward payer swaption is given by

$$\begin{aligned}\mathrm{PS}_0 &= \sum_{k=i+1}^{N} \delta_k B(0, T_k)\mathrm{E}^{\widetilde{P}_{T_{i+1}}}\left[(S(T_i, T_i, T_N) - K)_+\right] \\ &= C_{i,N}(0)\mathrm{E}^{\widetilde{P}_{T_{i+1}}}\left[(S(T_i, T_i, T_N) - K)_+\right].\end{aligned} \tag{26}$$

Eberlein and Raible (1999) and Raible (2000) propose a method for the evaluation of European stock options in a Lévy setting by using bilateral (or, two-sided) Laplace transforms. This approach is based on the observation that the pricing formula for European options can be represented as a convolution. Whence one can use the fact that the bilateral Laplace transform of a convolution is the product of the bilateral Laplace transforms of the factors (the latter transforms are usually known explicitly). Inversion of the bilateral Laplace transform then yields the option prices as a function of the current price of the underlying asset, and can be accomplished through the Fast Fourier Transform algorithm.

This method could also be employed for pricing the forward payer swaptions (26) as we shall shortly explain in the following. We concentrate on the purely discontinuous case ($c = 0$ in (24)) in view of applications using generalized hyperbolic Lévy processes. We consider the forward swap rate as given in (23). We define $X_t := \int_0^t \gamma(s, T_i) d\widetilde{L}_s^{T_{i+1}}$ so that $X_{T_i} = \ln\left(\frac{S(T_i,T_i,T_N)}{S(0,T_i,T_N)}\right)$. By defining $w(x, K) := (x - K)_+$, the payoff of the swaption is given by $w(S(T_i, T_i, T_N), K)$ and its price at time $t = 0$ by $\mathrm{E}^{\widetilde{P}_{T_{i+1}}}[w(S(T_i, T_i, T_N), K)]$. We consider the modified payoff $\widetilde{w}(x, K) := w(\mathrm{e}^{-x}, K)$.
Let $\zeta_i := -\ln S(0, T_i, T_N)$, then $S(T_i, T_i, T_N) = \mathrm{e}^{-\zeta_i + X_{T_i}}$. Furthermore, denote by $V(\zeta_i, K)$ the time zero price of the swaption, and let $L[\widetilde{w}]$ be the bilateral Laplace transform of $\widetilde{w}$:

$$L[\widetilde{w}](z) = \int_{-\infty}^{+\infty} \mathrm{e}^{-zx} \widetilde{w}(x) dx, \quad z = R + iu \in \mathbb{C}, \ \ R, u \in \mathbb{R}.$$

The price of the swaption at time zero can be written (apart from the discount factor) as a convolution of functions $\widetilde{w}(x)$ and $\rho(x)$, taken at the point ζ_i:

$$V(\zeta_i, K) = C_{i,N}(0)\mathrm{E}^{\widetilde{P}_{T_{i+1}}}[\widetilde{w}(\zeta_i - X_{T_i}, K)] = C_{i,N}(0) \int_{\mathbb{R}} \widetilde{w}(\zeta_i - x, K)\rho(x)(dx),$$

where ρ is the density function of X_{T_i}. As remarked above, the bilateral Laplace transform of a convolution equals the product of the bilateral Laplace transforms of the factors. Thus, we have that

$$L[V](R + iu) = C_{i,N}(0)L[\widetilde{w}](R + iu) \cdot L[\rho](R + iu). \tag{27}$$

As described in Raible (2000), we can invert the bilateral Laplace transform to obtain the swaption price V:

$$V(\zeta_i, K) = \frac{1}{2\pi i} \int_{R-i\infty}^{R+i\infty} \mathrm{e}^{\zeta_i z} L[V](z) dz = \frac{\mathrm{e}^{\zeta_i R}}{2\pi} \lim_{\substack{M \to \infty \\ N \to \infty}} \int_{-M}^{N} \mathrm{e}^{iu\zeta_i} L[V](R + iu) du. \tag{28}$$

Note that the identity $L[\rho](R + iu) = \chi(iR - u)$, where $\chi(iR - u) := \mathrm{E}^{\widetilde{P}_{T_{i+1}}}\left[\mathrm{e}^{i(iR-u)X_{T_i}}\right]$ is the extended characteristic function of X_{T_i}. By substituting (27) into (28) we obtain the swaption pricing formula

$$V(\zeta_i, K) = C_{i,N}(0) \frac{\mathrm{e}^{\zeta_i R}}{2\pi} \lim_{\substack{M \to \infty \\ N \to \infty}} \int_{-M}^{N} \mathrm{e}^{iu\zeta_i} L[\widetilde{w}](R + iu)\chi(iR - u) du. \tag{29}$$

According to Raible (2000) it is sufficient to consider the case where the strike price equals one, as
$$V(\zeta_i, K) = KV(\zeta_i + \ln K, 1).$$
The bilateral Laplace transform $L[\widetilde{w}]$ for $K = 1$ is given by $L[\widetilde{w}](z) = (z(z+1))^{-1}$, if $\mathbf{Re}z < -1$. We remark that the described approach can be used also for more complicated payoff functions as long as the payoff depends only on X_{T_i}. The characteristic function $\chi(u) := \mathrm{E}^{\widetilde{P}_{T_{i+1}}}\left[\mathrm{e}^{iuX_{T_i}}\right]$ can be determined more precisely once the distribution of L_1 is specified. Hence, we can calculate equation (29) numerically in an efficient way.

Acknowledgements

Jan Liinev gratefully acknowledges the financial support of the BOF-project 001104599 of the Ghent University, and of the European Community's Human Potential Programme under contract HPRN-CT-2000-00100, DYNSTOCH.

References

[1] Björk T., Di Masi G., Kabanov Y. and Runggaldier W. (1997): Towards a general theory of bond markets. *Finance and Stochastics*,**1**, 141-174.

[2] Björk T., Kabanov Y. and Runggaldier W. (1997): Bond market structure in the presence of marked point processes. *Mathematical Finance*, **7**, 211-239.

[3] Eberlein E. (2001): Application of generalized hyperbolic Lévy motions in finance. In: Barndorff-Nielsen O.E, Mikosch T. and Resnik S.I. (Eds.), *Lévy Processes: Theory and Applications*, Birkhäuser, 319-336.

[4] Eberlein E. and Özkan F. (2003): The Lévy Libor model. *FDM Preprint*, University of Freiburg.

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

[6] Glasserman P. and Kou S.G. (2003): The term structure of simple forward rates with jump risk. *Mathematical Finance*, **13**(3), 383-410.

[7] Heath D., Jarrow R. and Morton A. (1992): Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, **60**, 77-105.

[8] Hunt P.J. and Kennedy J.E. (2000): *Financial Derivatives in Theory and Practice*. Wiley Series in Probability and Statistics.

[9] Jacod J. and Shiryaev A.N. (1987): *Limit Theorems for Stochastic Processes*. Springer-Verlag.

[10] Jamshidian F. (1997): Libor and swap market models and measures. *Finance and Stochastics*, **4**, 293-330.

[11] Jamshidian F. (1999): Libor market model with semimartingales. *Working paper*, NetAnalytic Ltd.

[12] Liinev J. (2003): Topics in Mathematical Finance: from Interest Rate Models to Exotic Options. *Ph.D. thesis*, Ghent University.

[13] Musiela M. and Rutkowski M. (1997a): Continuous-time term structure models: Forward measure approach. *Finance and Stochastics*, **4**, 261-292.

[14] Musiela M. and Rutkowski M. (1997b): *Martingale Methods in Financial Modelling*. Springer-Verlag.

[15] Pelsser A. (2000): *Efficient Methods for Valuing Interest Rate Derivatives*. Springer-Verlag.

[16] Raible S. (2000): Lévy Processes in Finance: Theory, Numerics, and Empirical Facts. *Ph.D. thesis*, University of Freiburg.

[17] Rutkowski M. (1999): Models of forward Libor and swap rates. *Applied Mathematical Finance*, **6**, 1-32.

[18] Rutkowski M. (2001): Modelling of forward Libor and swap rates. Published in: Jouini E., Cvitanić J. and Musiela M. (Eds.) *Option Pricing, Interest Rates and Risk Management*, Handbooks in Mathematical Finance, Cambridge University Press, 336-395.

THE COMPOUND OPTION: AN OVERVIEW

Liesbeth Thomassen and Martine Van Wouwe

University of Antwerp, Faculty of Applied Economics, Antwerp, Belgium
Email: `liesbeth.thomassen.1@ua.ac.be`

Abstract

In this article we give an overview of the compound option theory and generalize this idea to the n-fold compound options. Furthermore, the use of this option valuation in the financial world and in other areas such as pharmaceutical R&D, is illustrated.

1. INTRODUCTION

1.1. Some notations

The following notations are introduced:

$$
\begin{aligned}
\overline{a}^{ij} &= (a_i, a_{i+1}, \ldots, a_j) \\
\overline{b}^{ij} &= (b_i, b_{i+1}, \ldots, b_j) \\
a_s &= b_s + \sigma\sqrt{t_s - t} \\
b_s &= \frac{\ln \frac{V}{V_s} + \left(r - \frac{1}{2}\sigma^2\right)(t_s - t)}{\sigma\sqrt{t_s - t}} \\
R_i^j &= \text{covariance matrix of } (X_i, X_{i+1}, \ldots, X_j) \\
& \quad cov(X_v, X_w) = \sqrt{\frac{t_v - t}{t_w - t}} \qquad v < w.
\end{aligned}
\tag{1}
$$

We also use the standard notations r for the risk-free interest rate, V for the value of the underlying asset and σ to denote its volatility. The value V of the asset underlying the considered options is supposed to follow a geometric Brownian motion unless mentioned otherwise. The dates t_i are the exercise dates of a compound option and later on in this article we will also use the notations K_i for the corresponding exercise prices.
If we now introduce functions $C_s(V, t)$, $\forall s = 1, \ldots, n$ as $(n - s + 1)$-fold compound options, it

is possible to compute the critical value $\overline{V}_s$ of V for which at time t_s the function C_{s+1} equals the exercise price K_s:

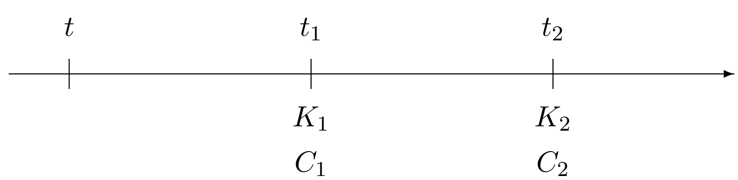
$$\overline{V}_s \quad \Rightarrow \quad C_{s+1}(\overline{V}_s, t_s) = K_s\,,$$

where for convenience $\overline{V}_n = K_n$.
Furthermore, we need the k-variate normal cumulative distribution function (CDF), which is given by

$$N_k(\overline{a}^{1k}; \overline{0}, R_1^k) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_k} \frac{1}{\sqrt{(2\pi)^k \cdot \det(R_1^k)}} \exp\left[-\frac{1}{2}\overline{x}^t (R_1^k)^{-1}\overline{x}\right] d\overline{x}\,,$$

with zero mean, covariance matrix R_1^k and with boundaries defined by the k-variate vector $\overline{a}^{1k}$.

1.2. The compound option



In introducing the compound option in 1979, Geske [4] wanted to value European call options with as underlying a European call option. He considered the European call option $C_1(V, t)$ with maturity date t_1 and strike price K_1 and as underlying asset a European call option $C_2(V, t)$ with maturity date t_2 and strike price K_2.
In assuming that the underlying asset V of the call C_2 follows a Brownian motion (as in the typical Black-Scholes setting), the value of the underlying call C_2 is known as a function of V and given by the well-known formula:

$$C_2(V, t) \quad = \quad V \cdot N_1(\overline{a}^{11}; \overline{0}, R_1^1) - K_2 \cdot e^{-r(t_2 - t)} \cdot N_1(\overline{b}^{11}; \overline{0}, R_1^1). \tag{2}$$

In deriving the value of the compound call C_1 by an analogous strategy but with an adapted boundary condition, he managed to prove a closed-form formula for C_1:

$$\begin{aligned} C_1(V, t) \quad = \quad & V \cdot N_2(\overline{a}^2; \overline{0}, R_2) - K_1 \cdot e^{-r(t_1 - t)} \cdot N_1(\overline{b}^1; \overline{0}, R_1) \\ & - K_2 \cdot e^{-r(t_2 - t)} \cdot N_2(\overline{b}^2; \overline{0}, R_2). \end{aligned} \tag{3}$$

Remark: the notations in (2) and (3) were introduced in section 1.1.

1.3. The n-fold compound option

t	t_1	t_2	t_3	$\ldots$	t_n
	K_1	K_2	K_3	$\ldots$	K_n
	C_1	C_2	C_3	$\ldots$	C_n

We generalize this idea to a composition of European call options, or roughly speaking an option on an option on an option on .... In proving some transformation theorems for the multivariate normal CDF, an analogous strategy makes it possible to value such an n-fold compound call. If we use the notations C_n for the usual Black-Scholes option, C_{n-1} for the 2-fold compound option with C_n as underlying option, ..., C_1 for the n-fold compound option, and if we suppose that all the values C_i are already defined for $i < n$, the resulting closed-form formula for C_1 is given by the following theorem:

Theorem 1.1 *If the $(n+1-i)$-fold compound call options C_i are known and defined by:*

$$C_i(V,t) \;=\; V \cdot N_{n+1-i}(\overline{a}^{in};\overline{0},R_i^n) - \sum_{j=i}^{n} K_j \cdot e^{-r(t_j-t)} \cdot N_{j+1-i}(\overline{b}^{ij};\overline{0},R_i^j),$$

the n-fold compound call option C_1 is given by:

$$C_1(V,t) \;=\; V \cdot N_n(\overline{a}^{1n};\overline{0},R_1^n) - \sum_{i=1}^{n} K_i \cdot e^{-r(t_i-t)} \cdot N_i(\overline{b}^{1i};\overline{0},R_1^i).$$

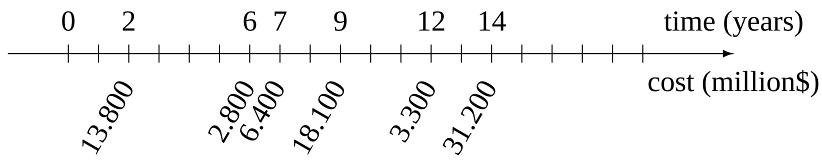
For a full proof of this theorem we refer to Thomassen and Van Wouwe [13].

2. PRACTICAL USE OF THE COMPOUND OPTION

Some practical applications of the n-fold compound option are discussed, as there are R&D developments, American options,...

2.1. R&D in the pharmaceutical world

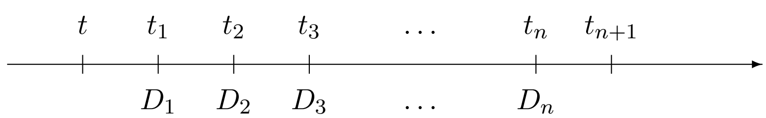
This first application is the development of a new drug, which typically evolves in 6 stages: discovery, preclinical testing, three clinical test phases (each time on larger test groups), FDA approval and finally the post-marketing testing. If now the new drug fails a test, an investor wants to have the opportunity to withdraw from the process. So the whole process can be seen as a 6-fold compound option because at the end of the subsequent phases, the investor has the possibility to leave in case of bad test results or to continue if not. The average drug testing process follows this scheme concerning time schedule and investment costs:



The evaluation of the corresponding 6-fold compound option leads to a resulting value of $27.500 million. The following conclusion can be drawn: considering an initial cost of $2.200 million to start the development, it is really worthwhile to invest in this R&D because the whole process adds value compared to the initial investment.
A detailed description of this application can be found in Cassimon et al. [2].

2.2. Unprotected American call option on stocks with discrete known dividends

The compound option theory can be used to derive closed-form formulas for an American call option with an underlying asset paying discrete known dividends. The initial valuation is performed by Roll, Geske and Whaley ([5], [6], [8] and [15]) for an option with at most 2 dividends.
A generalization of the formula toward an arbitrary amount of payment dates can be obtained as follows. Suppose we want to value an American call option with exercise price K and maturity date t_{n+1} on some underlying asset paying dividends D_i at intermediate dates t_i, $i = 1, 2, \ldots, n$.



Consider the following hedging portfolio P:

a) a long position on an American call option with maturity date t_{n+1} and exercise price K on a stock paying $n-1$ dividends ($D_2, \ldots, D_n$),

b) a long position on a European call option with exercise price $\overline{V} + D_1$ and exercise date $t_1 - \epsilon$ on a stock paying 1 dividend (D_1) during the life of the option,

c) a short position on a compound option, composed as follows: a European call option with exercise price $\overline{V} + D_1 - K$ and exercise date $t_1 - \epsilon$ with an underlying American call option as the one in a),

with $\overline{V}$ representing the critical value of V (ex-dividend) above which the American call option will be exercised at a time just prior to t_1. Clearly, both the portfolio P and the American call have the same value by the principle of no-arbitrage.
We suppose $\tilde{V}$ to follow a log-normal process instead of V, to avoid a positive probability of not being able to pay the dividends in the future:

$$\tilde{V} = V - \sum_{i=j+1}^{n} D_i \, e^{-r(t_i - t)}, \qquad t_j^+ \leq t \leq t_{j+1}^-, \qquad j = 0, 1, \ldots, n.$$

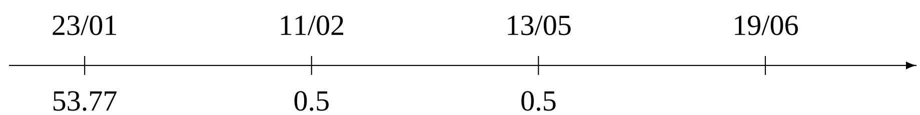
Using the compound option theory together with the induction principle, a closed-form formula for the American call option is obtained:

$$
\begin{aligned}
C_{A,n}(\tilde{V},t) &= \tilde{V}\cdot\left[1 - N_{n+1}(-\overline{a}^{1,n+1};\overline{0},R_1^{n+1})\right] \\
&\quad -K\sum_{i=2}^{n+1} e^{-r(t_i-t)}\,N_i(-\overline{b}^{1i*};\overline{0},R_1^{i*}) - K\,e^{-r(t_1-t)}\,N_1(b_1;\overline{0},R_1^1) \\
&\quad +\sum_{i=2}^{n} D_i\,e^{-r(t_i-t)}\sum_{j=2}^{i} N_j(-\overline{b}^{1j*};\overline{0},R_1^{j*}) \\
&\quad +\sum_{i=1}^{n} D_i\,e^{-r(t_i-t)}\,N_1(b_1;\overline{0},R_1^1)
\end{aligned}
$$

with

$$
\begin{aligned}
\overline{b}^{1i*} &= (b_1,b_2,\ldots,b_{i-1},-b_i) \\
R_1^{i*} &= \text{the covariance matrix of } (X_1,X_2,\ldots,X_{i-1},-X_i).
\end{aligned}
$$

Example: In January 2004, American call options could be bought on assets of GM. The value of such an asset was \$53.77 on the 23^{th} of January and the asset would pay \$0.5 dividend on the 11^{th} of February and on the 13^{th} of May. The option matures the 19^{th} of June.



We use the Fortran code MVNDST by Genz [3] to evaluate the multivariate normal CDF's in the American option formula and obtain the following comparison between the real market prices and the theoretical values of our model:

K	market	$\overline{V}_1$	$\overline{V}_2$	model	Eur.
45	8.90	49.03	47.31	8.89	8.70
50	5.00	55.18	52.88	4.82	4.74
55	2.20	61.45	58.48	2.13	2.09
60	0.80	67.84	64.13	0.75	0.74
65	0.25	74.36	69.82	0.22	0.21

We mentioned the critical prices $\overline{V}_1$ and $\overline{V}_2$ above which the value V should rise at date t_1 respectively t_2, before early exercise becomes profitable at these moments.
In the last column, the prices of a corresponding European call option are given. Clearly, these prices fall below the American option prices, as expected theoretically.

3. DECOMPOSITION OF THE N-FOLD COMPOUND OPTION

The n-fold compound option was initially defined as a 1-fold on an (n-1)-fold. Further research about the sensitivity of the n-fold toward for instance the position of the intermediate dates and the

related exercise prices, urged for more general decompositions of the n-fold.
Therefore we proved that an n-fold compound option can be constructed as an (n-k)-fold on a k-fold:

Theorem 3.1 **First PDE**: *C_1 is the solution of*

$$\frac{\partial C_1}{\partial t} = r\,C_1 - r\,V\,\frac{\partial C_1}{\partial V} - \frac{1}{2}\,\sigma_V^2\,V^2\,\frac{\partial^2 C_1}{\partial V^2}. \tag{4}$$

Second PDE: *C_1 is the solution of*

$$\frac{\partial C_1}{\partial t} = r\,C_1 - r\,C_{n-k+1}\,\frac{\partial C_1}{\partial C_{n-k+1}} - \frac{1}{2}\,\sigma_{n-k+1}^2\,C_{n-k+1}^2\,\frac{\partial^2 C_1}{\partial C_{n-k+1}^2}, \tag{5}$$

with C_{n-k+1} the underlying asset, satisfying itself a similar PDE

$$\frac{\partial C_{n-k+1}}{\partial t} = r\,C_{n-k+1} - r\,V\,\frac{\partial C_{n-k+1}}{\partial V} - \frac{1}{2}\,\sigma_V^2\,V^2\,\frac{\partial^2 C_{n-k+1}}{\partial V^2}\,. \tag{6}$$

PDE (5) can be rewritten into PDE (4) with the specific boundary conditions.

A detailed proof of the theorem can be found in Thomassen, Van Casteren and Van Wouwe [12].
Another advantage of this theorem is that it permits a controlling mechanism for the numerical results. If for instance an n-fold is calculated in several ways, the same numerical value should be obtained for the n-fold. Suppose a virtual case where $r = 0.05$, $\sigma = 0.2$, $V = 90$. Consider a 4-fold compound option with exercise dates $2, 5, 9, 10$ and each exercise price equal to 1. Valuing the 4-fold at time $t = 0$ results in:

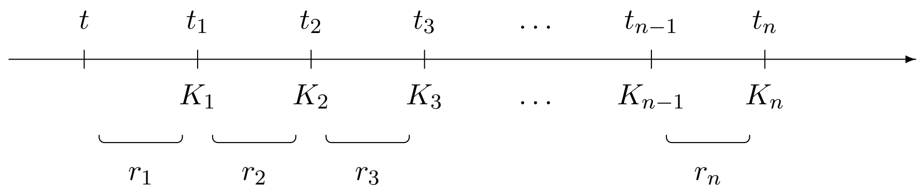
4-fold option as:	value
1-fold on 3-fold	87.07220298755823
2-fold on 2-fold	87.07220298755823
3-fold on 1-fold	87.07220298755823
1-fold on 1-fold on 1-fold on 1-fold	87.07220298755824
4-fold	87.07220298755823

4. GENERALIZATION OF THE MODEL

The setting of n-fold compound options is performed in the well-known world of Black and Scholes. However, in looking at the practical use of n-fold compound options, it is clear that it can invoke long-term contracts, so that the assumption of a constant interest rate over the whole period seems a bit unrealistic.

4.1. Discrete change in interest rate

As a first generalization, we allow the interest rate to exhibit discrete changes over each interval between two subsequent exercise dates:



Again under this assumption a closed-form formula can be obtained for the value of an n-fold compound call option:

$$\begin{aligned} C_1(V,t) &= V \cdot N_n(\overline{a}^{1n};\overline{0},R_1^n) \\ &\quad - \sum_{i=1}^{n} K_i\, e^{-r_1(t_1-t)} \prod_{j=2}^{i} e^{-r_j(t_j-t_{j-1})}\, N_i(\overline{b}^{1i};\overline{0},R_1^i) \end{aligned}$$

where the vectors $\overline{b}^{1i}$ now are given by:

$$\begin{aligned} b_j &= \frac{\ln \frac{V}{V_j} + (r_j - \frac{1}{2}\sigma^2)(t_j - t)}{\sigma\sqrt{t_j - t}} + S_{1j} + \sum_{k=2}^{j-1} S_{kj} \left(\frac{t_k - t_{k-1}}{t_k - t}\right) \\ S_{kj} &= (r_k - r_j)\cdot\frac{t_k - t}{\sigma\sqrt{t_j - t}} \end{aligned}$$

and where $\overline{a}^{1n}$ still is defined according to equation (1).

4.2. Continuous interest rate

This subsection is based on the work by Miltersen, Sandmann and Sondermann [7], who obtained a closed-form formula for a European call option in the setting of a stochastic interest rate.
They supposed the simple forward rates f to follow a log-normal distribution, to avoid both the problem of possibly negative interest rates and exploding interest rates:

$$df(\cdot,T,\alpha)_t = \mu(t,T,\alpha)\cdot f(t,T,\alpha)dt + \gamma(t,T,\alpha)\cdot f(t,T,\alpha)dW_t$$

In this model $P(t,T+\alpha)$, the value for a zero-coupon bond maturing at time $T+\alpha$, and $F(t,T,\alpha)$, a forward contract (to buy at time T a zero-coupon maturing at $T+\alpha$), are:

$$\begin{aligned} P(t,T+\alpha) &= P(t,T)\frac{1}{1+\alpha\cdot f(t,T,\alpha)} \\ F(t,T,\alpha) &= \frac{P(t,T+\alpha)}{P(t,T)} = \frac{1}{1+\alpha\cdot f(t,T,\alpha)} \end{aligned}$$

Miltersen, Sandmann and Sondermann [7] found the following closed-form formula for a European call option with maturity date T, exercise price K and with as underlying asset $P(t,T+\alpha)$ a zero-coupon bond maturing at $T+\alpha$:

$$\begin{aligned} C\Big(P(t,T+\alpha),t\Big) &= P(t,T+\alpha)\cdot(1-K)\cdot N_1(a_1;\overline{0},R_1^1) \\ &\quad -K\Big(P(t,T) - P(t,T+\alpha)\Big)\, N_1(b_1;\overline{0},R_1^1) \end{aligned}$$

where:

$$
\begin{aligned}
a_1 &= b_1 + \sqrt{s_1} \\
b_1 &= \frac{1}{\sqrt{s_1}} \left[\ln \left| \frac{P(t,T+\alpha)\cdot(1-K)}{K\cdot\Big(P(t,T)-P(t,T+\alpha)\Big)} \right| - \frac{s_1}{2} \right] \\
s_1 &= \int_t^T \gamma^2(u,T,\alpha)du
\end{aligned}
$$

Again we are able to generalize this expression to an n-fold compound call with exercise dates t_i, exercise prices K_i for $i = 1,\ldots,n$, and a zero-coupon bond $P(t, t_n + \alpha_n)$ as underlying.
The closed-form formula for the n-fold C_1 in terms of its forward value $\hat{C}_1$ is:

$$
\begin{aligned}
\hat{C}_1 &= F\cdot N_n\Big(\overline{a}^{1n};\overline{0},G_n\Big)\cdot\prod_{i=1}^{n-1}P(t_i,t_{i+1}) \\
&\quad -\sum_{i=1}^{n}K_i\cdot F\cdot N_i\Big(\overline{a}^{1i};\overline{0},G_i\Big)\prod_{j=1}^{i-1}P(t_j,t_{j+1}) \\
&\quad -\sum_{i=1}^{n}K_i\cdot(1-F)\cdot N_i\Big(\overline{b}^{1i};\overline{0},G_i\Big)\prod_{j=1}^{i-1}P(t_j,t_{j+1})
\end{aligned}
$$

where the relation between the n-fold compound option and its forward value is given by:

$$
C_1 = P(t,t_1)\cdot\hat{C}_1
$$

and

$$
\begin{aligned}
F &= F(t,t_n,\alpha_n) \quad \text{(short notation)} \\
a_j &= b_j + \sqrt{s_j} \\
b_j &= \frac{1}{\sqrt{s_1}}\left[\ln\left|\frac{F\cdot(1-F_j)}{F_j\cdot(1-F)}\right| - \frac{s_j}{2}\right] \\
s_j &= \int_t^{t_j}\gamma^2(u,t_n,\alpha_n)du \\
F_j &: \text{ solution of } P(t_j,t_{j+1})\cdot\hat{C}_{j+1}(F(t,t_n,\alpha_n),t_j) = K_j \\
G_k &= \left(\sqrt{\frac{s_i}{s_j}}\right)_{\substack{i=1,\ldots,j\\ j=1,\ldots,k}} \qquad \text{symmetric}
\end{aligned}
$$

5. FUTURE RESEARCH

Because a lot of real life processes are compounded, such as investment plans, R&D developments, ..., it is worthwhile not only to value such processes, but also to value the possibility of choosing

at certain intermediate dates between the continuation of the process or ending the process. This is a valuation strategy where n-fold compound options are needed.
Of course, a lot of research concerning valuation of such choices, or concerning the possible stretching of conditions in our model, still has to be performed. The last subsection is only a first step in the relaxation of conditions. It shows that it is possible to value n-fold compound options in a stochastic interest rate setting.

References

[1] Black F. and Scholes M. (1973): The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, **81**(3), 637-654.

[2] Cassimon D., Engelen P.J., Thomassen L. and Van Wouwe M. (2004): The valuation of a NDA Using a 6-fold Compound Option. *Research Policy*, **33**(1), 41-51.

[3] Genz A. (1992): Numerical Computation of Multivariate Normal Probabilities. *Journal of Computational and Graphical Statistics*, **1**, 141-149.

[4] Geske R. (1979): The Valuation of Compound Options. *Journal of Financial Economics*, **7**, 63-81.

[5] Geske R. (1979): A Note on an Analytical Valuation Formula for Unprotected American Call Options on Stocks with Known Dividends. *Journal of Financial Economics*, **7**(4), 375-380.

[6] Geske R. (1981): Comments on Whaley's Note. *Journal of Financial Economics*, **9**(2), 213-215.

[7] Miltersen K.R., Sandmann K. and Sondermann D. (1997): Closed Form Solutions for Term Structure Derivatives with Log-Normal Interest Rates, *The Journal of Finance*, **52**(1), 409-430.

[8] Roll R. (1977): An Analytic Valuation Formula for Unprotected American Call Options on Stocks with Known Dividends. *Journal of Financial Economics*, **5**(2), 251-258.

[9] Schervish M.J. (1984): Multivariate Normal Probabilities with Error Bound. *Journal of the Royal Statistical Society*, Series C, **33**, 81-94.

[10] Thomassen L. and Van Wouwe M. (2001): The n-fold Compound Option. *Research Paper 2001-041*, UA Faculty of Applied Economics, Antwerp.

[11] Thomassen L. and Van Wouwe M. (2002): A sensitivity analysis for the n-fold compound option. *Research Paper 2002-014*, UA Faculty of Applied Economics, Antwerp.

[12] Thomassen L., Van Casteren J. and Van Wouwe M. (2002): Decomposition of the n-fold Compound Option. *Research Paper 2002-040*, UA Faculty of Applied Economics, Antwerp.

[13] Thomassen L. and Van Wouwe M.: The n-fold Compound Option: Theory and Sensitivity, forthcoming.

[14] Thomassen L. and Van Wouwe M.: The influence of a stochastic interest rate on the n-fold compound option. *In: Theory and Applications of Recent Robust Methods*, Hubert M., Pison G., Struyf A. and Van Aelst S. (Eds.). Series: Statistics for Industry and Technology, Birkhauser, Basel, to appear.

[15] Whaley R.E. (1981): On the Valuation of American Call Options on Stocks with Known Dividends. *Journal of Financial Economics*, **9**(2), 207-211.

FINANCIAL CHALLENGES IN POWER MARKETS

Michel Verschuere

Luminus corporation, Hasselt, Belgium and Institute for Financial and Actuarial Mathematics, TU Vienna, Austria.
Email: `michel@fam.tuwien.ac.at`

Abstract

We review the ongoing deregulation process in power markets around the globe. We point out where financial challenges remain in electricity price risk management and we revisit existing literature on the subject.

1. INTRODUCTION

In recent years, we have witnessed a worldwide tendency towards deregulation in network industries and energy markets. The main driving force behind the deregularization is a quest for increased competition with the ultimate aim to reduce prices for end-users. The UK was the first state to create the legal framework for an open power market in the form of the 1989 Electricity Act. In mainland Europe, the Nordic countries Denmark, Finland, Norway and Sweden followed the UK's example when creating the Nordic exchange area in 1998. All EU member states are committed to the opening of their domestic power markets by 2007 and opening has been achieved to a varying extent as we write.

The deregularization of domestic electricity markets involves the vertical separation of the once fully state-controlled power sector. The generation-, transport-, distribution- and supply building blocks were disentangled and private companies now compete with one-another within most of these branches. A state monopoly usually remains in the transport segment, as it requires the sort of large scale investment in infrastructure that renders it less competitive. Power exchanges or 'pools' were then created, where electrical energy is traded by the megawatthour (MWh). Next to physical and financial trading floors, they sometimes act as a clearing house for OTC supply agreements. Nowadays, electrical energy is increasingly traded as a standard commodity, despite its unique properties.

Electrical energy is generated out of basic energy sources such as fossile fuels, wind-, water- or nuclear forces. The produced power is then transported over high voltage lines before finally being supplied to end-users. The main difference between electrical energy and other commodities is the property that it cannot be economically stored once generated. Only countries with hydroelectric capabilities have a means to store generated power for later use, albeit indirectly. In addition, the unstoreable nature of electrical energy has dramatic consequences for its price behaviour as we shall see later.

This paper is structured as follows: We first discuss the stylized facts of power spot prices and we refer to the known mathematical models to capture them. Section three deals with the nature and properties of forward prices. Although futures and forwards are principally a form of derivatives contract, they should be distinguished from the standard power options and more exotic types treated in section four. Pricing and hedging of electricity derivatives is severely restricted by the unstoreable character of the underlying. Nonstandard approaches are therefore required and we discuss a few of them before concluding this note in section five. Our objective here is to provide an overview of topics that are of interest to both financial practitioners and researchers. The bibliography is representative for state-of-the-art economical- and mathematical research in power markets.

2. POWER SPOT PRICES

2.1. Stylized facts

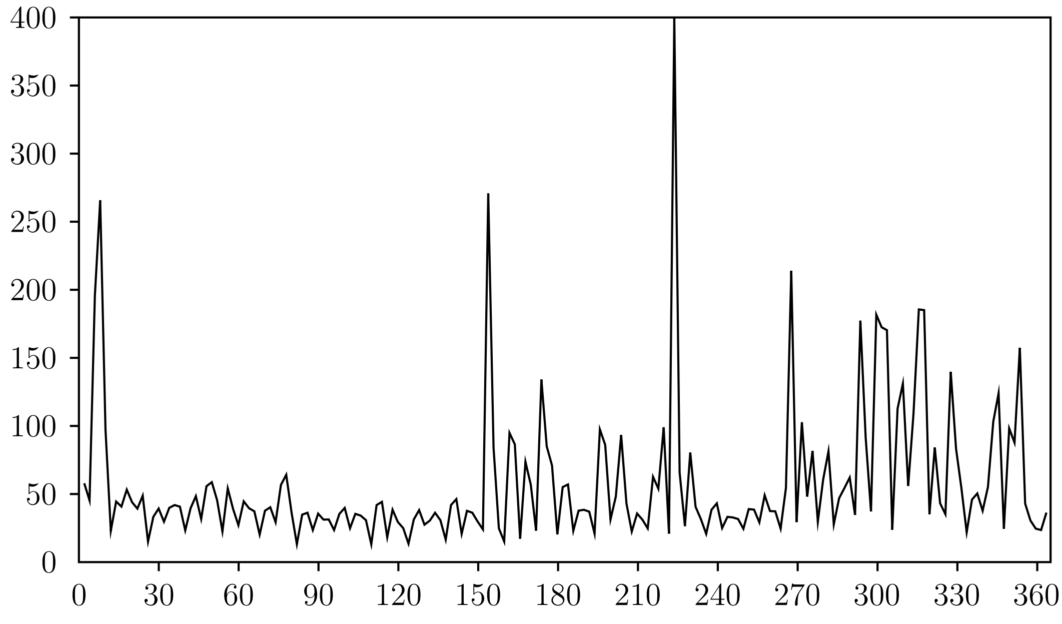


Figure 1: Daily peakload (07-23h) power spot prices at the APX, 2003 (Euro/MWh). Price ticks above 400 EUR/MWh have been skipped. [1].

Figure 1 displays peakload spot prices recorded from the Dutch APX power auction over the year 2003. This chart confronts us with generic behaviour of prices for the physical delivery of one MWh of electrical energy over each of the 24 hours of the following day. Most power exchanges provide an electronic 'trading floor' for bulk delivery on the next day in the form of an auction that is often referred to as a 'one-day-ahead market'. Both generators and suppliers submit their respective supply- and demand bids for all 24 hourly intervals of the next day. An automated trading system then establishes the equilibrium 'system prices', after which all participants are informed on both prices and delivery schedules. Power prices corresponding to different hourly intervals usually differ due to changing demand or 'load' patterns.

The most important stylized facts observed in power spot prices are *strong seasonality*, *mean reversion* and so-called price *spikes*. The *causality* of price fluctuations is worth mentioning too, meaning that that they reflect instantaneous supply and demand levels as an immediate consequence of the unstoreable nature of electrical energy. We discuss these important stylized facts in the remainder of this section.

The *strong seasonality* observed in one-day-ahead prices for power occurs on a variety of time scales. There are periodic patterns on an hourly, daily, weekly and seasonal basis that go hand in hand with similar fluctuations in the demand for this commodity. Seasonal patterns are typically observed in prices for goods that suffer from storage constraints, think about agricultural products such as wheat for instance. The effect is further enhanced for products that are difficult or expensive to stockpile and it meets an extreme end in case of power. Electricity demand is usually higher during the day and consumption drops over weekends or on holidays. The load increases in winter and summer, respectively due to heating and air conditioning. These 'foreseeable' demand variations reflect themselves in prices as the latest generation plant employed to meet demand levels delivers at the highest marginal cost. This leads to an aggregated supply curve that is upward sloping.

The *mean reverting* feature of power spot prices can be explained from the generation process. Most power is generated by combusting fossile fuels such as coal, natural gas, oil or through nuclear fission. Renewable energy sources such as wind-, water- and biothermal power are a relatively new phenomenon still, while they are sometimes unavailable for geographic reasons. Consequently, marginal cost levels fluctuate around long term averages set by prices for these more basic fuel sources, augmented with plant-managing fees and profit margins. The bulk of the annual seasonality in power prices is in most cases an immediate consequence of similar periodic patterns in basic fuel costs.

Sudden, short-lived but dramatic price rises are the most striking feature of power spot prices. In early August 2003, prices of up to 3000 Euro/MWh were recorded at the Dutch APX exchange, about a 100 times the average as figure 1 indicates. Needless to say that such *price spikes* can severely damage the financial health of market participants and it drives their quest for methods to mitigate this risk. Price spikes may be induced by transmission failures, breakdowns at generation plants or simply by extreme weather conditions boosting load levels further up in times of peak demand. Clearly, such events are hard to predict long enough in advance.

Consider a power market where a 100 MW plant suddenly breaks down. Repairs are estimated to last over the next 24 hours and providers must search the one-day-ahead market for 100 MW of replacement capacity as they are obliged to meet customer needs. In extreme situations such as peak load and hot periods, the spot market may turn out too illiquid to deal with this unforeseen demand level and prices skyrocket as a consequence. Only in markets with hydroelectric generation capabilities one may encounter smoother price behaviour and the Scandinavian market here serves as a clear cut example.

2.2. Mathematical models

A key reference on mathematical spot price models is the work by Lucia and Schwartz [15]. This paper treats spot prices arising at the Nordic Exchange and discusses the performance of one- and two factor models for capturing all main trends. Other references are [5] and the regime switching model introduced in [10] that assumes two different market states in one of which price spikes arise. The regime switching model was recently revisited in [18] where an insurance premium formula was derived to cover spot price risk. Numerous attempts intend to capture power spot price behaviour in micro-economically inspired models and bottom up models form the most sophisticated example. An economically intuitive and appealing attempt was made by Barlow in [2]. Spot prices are there obtained from a nonlinear mapping of a one-factor mean reverting diffusion process, reflecting the inverse mapping of exogenous given demand by a stylized supply curve. A second micro-economical approach is described in the paper by Elliott *et al.* in [7], where price spikes arise as a consequence of large plants going offline. Their model was inspired by the Alberta power market where only fourteen different plants are present. Further attempts involve the modelling of the auction pricing that is behind price formation in power markets and the reader is referred to [9] and references therein. The number of statistical surveys of power spot price data is limited, but the work of Weron [23] deserves mentioning here.

3. POWER FORWARD PRICES

3.1. Stylized facts

Many power exchanges trade forward contracts as a primary form of electricity derivative. These contracts are highly standardized, for instance entailing the financial supply of a constant flow of 1 MWh per hour within a delivery period specified in the contract. Their counterparts for physical delivery trade in the OTC market and exchanges like NordPool sometimes also act as a clearing house for such agreements. At NordPool, delivery periods vary from days over weeks, blocks (four weeks) and seasons, up to entire years. Contracts for delivery over yearly periods become available about four years in advance and gradually decompose into contracts with shorter delivery windows. The ‘forward cascade’ is the set of rules maintained by the exchange to decompose the forward contracts into contracts with shorter delivery periods. The cascading occurs in a fully deterministic fashion until the shortest delivery period is met.

The daily, weekly and block financial contracts traded at NordPool are futures contracts while the remaining instruments are of the pure forward type. In both cases, gains and losses are settled through a margin account, but in case of forwards, they are accumulated up to the instant of delivery. A futures contract can be entered into at no cost, but eventual gains are settled in a daily ‘marked-to-market’ procedure involving the margin account.

Consider an example futures contract that delivers 1 MWh per hour over a one week period, corresponding to $(7 \times 24) = 168$ individual hours. Todays price equals 30.00 EUR/MWh and one such contract is entered into. In case tomorrows closing price for the same contract equals 31.00 EUR/MWh, the exchange will increase the margin account by $168 \times (31.00 - 30.00) = 168$ EUR. Similarly, when prices drop to 29 EUR/MWh, the exchange will kindly pass a 168 EUR margin call per contract to all holders of a long futures position. In this way, the futures trader will have received the difference (delivery price $-$ entered price) by the time the delivery takes place, just as with a forward contract. The only difference is that the futures settlement occurs over the entire period before delivery, involving the interest-bearing margin account. Apart from this 'interest rate convexity', futures- and forward positions are completely identical financially. Both futures and forwards are eventually settled on an 'ex-post' basis: Positions are gradually cleared over the delivery period. In the above example, the holder of the futures contract receives 24 times the difference (average daily price $-$ 30) EUR/MWh, every day within the delivery week.

Futures and forward contracts are de facto power derivatives because one-day-ahead (spot) prices remain the basic underlying. Forward positions prove highly valuable in the volatile physical markets, allowing for price risk to be spread over longer periods. Forward markets also prove convenient in long term decision making, as prices indicate future cost levels as anticipated by the aggregated market.

Consider a power market where a governmental decision is made to close an important nuclear plant in five years. The plant was known to be a reliable source of cheap baseload electrical energy and as soon as the decision is made public, power forward prices for delivery within five years or later start to rise. These higher prices should stimulate market participants to invest in fresh production capacity that is to replace the Nuclear plant. In a regulated environment capacity planning used to be a public matter but a forward market naturally completes the feedback loop in any liberalized market.

The most noteworthy feature of power forward prices is the *anomalous behaviour* encountered as maturity closes in. This property groups two different effects that strengthen as time-to-maturity decreases: one is the sharp increase in volatility also observed in other commodity futures markets where it is known as Samuelson effect. The second one is the appearance of an unusual stochastic drift that becomes stronger near expiry.

These facts have an important impact on mathematical modelling attempts. Futures prices converge against the spot price level at the instant of delivery, since positions in one such futures contract or 1 MWh of time-T spot power are financially equivalent at time T, thus in every continuous model one has

$$\lim_{t \uparrow T} F(t,T) = X_T, \tag{1}$$

where $F(t,T)$ is the forward price of one MWh at time t for delivery at time T and X_T is the spot price of one MWh at time T. As the latter is known to be very volatile, one has an intuitive explanation for the anomalous effect mentioned, occurring for times t close to T.

The equality in (1) may be violated in some special circumstances. In case physical power can not be delivered due to transmission failures, the futures- and spot price levels may decouple, see for instance [21]. This of course is an argument against the use of continuous models. Furthermore, relation (1) is usually blurred by the forward cascade as futures prices merely reflect 'rational expectations' of spot price levels over the delivery period. But generally any mathematical attempt to model the $F(t,T)$ futures prices should yield a spot price model with the accustomed properties through (1). An interesting approach in this sense can be retrieved in [22] and it was inspired upon the spot price model introduced by Barlow in [2].

3.2. Mathematical models

Reference [15] is a good start for our literature survey on power forward prices. The mathematical spot price models treated there imply futures and forward dynamics on no arbitrage grounds. Such an attempt can only be approximately valid, as the unstoreable nature of electrical energy seriously limits dynamic hedging strategies. In an empirical study of power forward curves reported in [13], it was found that prices referring to different instants of delivery vary in less dependent ways compared to other commodities. Because of the severe storage constraints, such contracts become completely different financial vehicles. The anomalous effect was first described in [19] and a micro-economical model was introduced to provide a possible explanation. The anomalous effect was implicitly treated in [3], where an increasing volatility structure was employed to price vanilla futures derivatives in the Nordic market yet also by questionable no-arbitrage methods. For a statistical survey of power forward prices, we refer to [14], which provides a detailed analysis of American PJM prices. One of the main challenges in power markets remains to define accurate models for the term structure dynamics that also provide satisfactory spot price behaviour through the limit (1). As far as we know, only one such an 'integrated' model was introduced sofar in [4] apart from [22] and both required very sophisticated approaches.

4. ELECTRICITY DERIVATIVES

Derivatives contracts seem essential risk managing tools in the volatile spot markets. In power markets, such financial vehicles appear in a variety of different forms and this section contains a comprehensive classification of these types, together with some comments on their valuation.

4.1. Overview

A first derivatives contract is the *futures option*, i.e. a European Call or Put option that is written on the price for a futures or forward contract at a given strike. Such options are openly quoted at NordPool and they usually expire shortly before financial delivery of the underlying commences. At NordPool, trade in these contracts is often illiquid, in contrast to the higher daily trading volumes for their underlying. The pricing and hedging of futures options requires a model for the price behaviour of futures contracts that should capture the anomalous behaviour mentioned in section three. In the simplest case, one employs a geometric Brownian motion process with a time-dependent volatility, leading to a valuation formula of the Black-Scholes type. Such an attempt was suggested in [3], [19] and [20]. The derivative can be replicated by a portfolio consisting of a position in the underlying forward contract and a cash account.

It turns out that most power derivatives are OTC agreements. Their prices are not openly quoted and they are often matched to the buyers' needs. Examples of such contracts are of the *Asian type*, *cross-commodity derivatives*, *virtual power plants*, *swing options* and ordinary *insurances*.

The *Asian type options* are bilateral agreements that entail a settlement at the end of a delivery period, according to the average spot price registered in it. Let $X_t;\ t \geq 0$ represent the time-t spot price for power, delivered over the hour following t. An asian option with strike K EUR/MWh over the time window $[T_1, T_2]$ is a time-T_2 contingent claim with payoff $\varphi(T_2)$ given by

$$\varphi(T_2) = \max\{\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} X_u \, du - K, 0\} =: (\frac{1}{T_2 - T_1} \int_{T_1}^{T_2} X_u \, du - K)^+.$$

For most choices of the price process $\{X_t : \ t \geq 0\}$, the pricing and hedging question for Asians proves a very complicated one. It is challenging to try to understand how such contingent claims could 'best' be replicated (in least-square sense for instance), by means of widely available futures contracts. An approximate valuation technique was introduced in [3] and to-date and to our knowledge there is no alternative answer.

Spark spread options are tailored to the peculiarities of the power market. They are a form of *cross commodity derivative* based on the fact that electrical energy is produced out of more basic sources such as natural gas for instance. In such a gas-fired power market, extra power can be delivered at short notice and plant ramp-up times are often negligible. The spark spread is there defined as the power spot price minus the gas spot price times a conversion factor. Gas prices $Y_t; \ t \geq 0$ are quoted in EUR/Btu (British thermal units) and the conversion factor H is plant-specific. It tells us that H British thermal units of natural gas must be feeded to the plant in order to produce 1 MWh of electrical energy. The spark spread option is often a European type Call on the time-T value of the spark spread, i.e. it has a payoff $\varphi(T)$:

$$\varphi(T) = (X_T - H \cdot Y_T)^+,$$

where X_t is the time-t spot price for electrical power in EUR/MWh. Spark spread options can be replicated by a portfolio consisting of both gas- and power futures maturing at time T. However, such hedging attempts are jeopardized by the presence of the forward cascade that renders the required futures prices invisible up to shortly before time T. The reader is referred to [6] for additional discussion and quantitative analysis.

Power *swing options* and *virtual power plants* are financial contracts whose payoffs mimic the characteristics of a real power plant. The owner of a virtual power plant can access power at a predetermined unit price k, up to a specified upper bound p_{up} MW over the period of time $[T_1, T_2]$. He thus fictively possesses a plant with p_{up} as its maximum output that produces at the marginal cost k. In case of swing options, there is an additional lower bound $0 \leq p_{low} \leq p_{up}$ to the load pattern $p(t)$, i.e. the buyer of the contract is forced to accept at least p_{low} MW at all times. Furthermore, the total amount of purchased power is limited by

$$e_{low} \leq \int_{T_1}^{T_2} p(u) \, du \leq e_{up},$$

with $0 \leq e_{low} \leq e_{up}$. For a discussion on the partial replication of swing options with basic power derivatives, the reader is referred to Keppo [11] and references therein.

Insurance contracts are a final class of risk managing tools available to power market participants. One can think of circumstances where active hedging of spot price risk is either unfavourable or impossible, while some of that risk is passed on to end-users. Intermediate consumers of electrical energy may then come to bear considerable financial risk against which they want to be insured. Such contracts can be tailored to the buyers' needs and their valuation becomes a pure actuarial matter in the absence of any dynamic hedging attempts. One premium formula for European Call options written on power spot prices was inspired by the work [10] and reported on in [18]. Alternative premium principles can be of use here and we refer to the excellent reference [12] for a treatise on them.

4.2. Hedging aspects

There are basically two ways to hedge power derivatives: either directly by *operating a production plant*, see [8], or indirectly using a *futures market*. Let us illustrate both methods by an example situation.

Consider a European Call option, written on the power spot price process $\{X_t : \ t \geq 0\}$, expiring at time T against the strike K EUR/MWh. The time-T payoff for this contingent claim thus reads $\varphi(T)$, with

$$\varphi(T) = \max\{X_T - K, 0\} := (X_T - K)^+. \tag{2}$$

The writer of the derivatives contract happens to own a plant that produces electrical energy at marginal cost k EUR/MWh and we shall neglect fixed costs and ramp-up periods for a moment. Ownership of the plant is financially equivalent to a long position in the real option $\eta(t)$, with

$$\eta(t) = (X_t - k)^+,$$

as the plant will only go online in case the spot price X_t is above the marginal cost level k. In case $K = k$, the plant 'physically' replicates the European Call option. More generally, the wealth of such a plant owner that is short one Call is given by the difference $\eta(T) - \varphi(T)$, which is positive provided $K \geq k$, i.e. it is favourable to produce power at marginal costs below the strike price. Clearly, the time-t option value becomes dependent on the plants' characteristics through the marginal cost level k and Call premia will not be indifferent to it either. The production- based valuation of power derivatives therefore reduces to an optimization problem for the dispatch profile of the plant. Both historical price- and load levels and plant characteristics play an important role in this pricing stage. Such studies fit into the field of operations research rather than within financial mathematics, see the work by Hinz in [9] and references therein.

A liquid *futures market* provides promising financial opportunities to hedge power spot derivatives. The key observation here is the limit (1), expressing that futures prices converge towards spot price levels at delivery, provided basis risk is neglected. Let $F(t, T)$ denote the time-t price for 1 MWh of electrical energy delivered at T. At time T, the payoff for a European Call $\varphi(T)$ in (2) on the spot price becomes equivalent to

$$\varphi(T) = \lim_{t \uparrow T} (F(t, T) - K)^+,$$

and this identity suggest that time-T spot derivatives can be replicated by a portfolio consisting of a time-T futures position.

To fully exploit futures hedging of spot derivatives one needs term structure models for forward prices that yield consistent spot prices in the limit (1) and there still is a lack of satisfactory results at this point. Futures hedging is further limited by the presence of a forward cascade, as prices for the precise futures contract underlying the spot derivative may remain hidden in the market until shortly before expiry.

5. CONCLUSIONS AND OUTLOOK

We gave an overview of current financial research in deregulated power markets. Many mathematical challenges remain in this rapidly expanding field. There is a growing need for good term structure models that capture both typical forward price behaviour and the main stylized facts for power spot prices. Such mathematical models are a key requirement for pricing many different types of derivatives. It looks like this futures hedging is the only way in which these contingent claims can be hedged fully financially. A second method involves the use of physical production capacity that demands for deeper commitment to the power market. Market players that do not have any production capabilities must rely upon either futures contracts or energy insurances to mitigate combined price and volume risk.

We included many references to earlier work throughout the text, such as to give a comprehensive as well as up to date overview of the field. We hope that the present paper may prove useful to both theorists and practitioners of power risk management.

Acknowledgement

I am indebted to Dipl.-Math. Lutz v. Grafenstein (DFG-Graduiertenkolleg 251, TU Berlin) for proofreading this manuscript.

References

[1] The Amsterdam power exchange (2003): `http://www.apx.nl`.

[2] Barlow M. (2004): A diffusion model for electricity prices. *Mathematical Finance*, **12**, 1-31.

[3] Bjerksund P., Rasmussen H. and Stensland G. (2000): Valuation and risk management in the Norwegian electricity market. *Discussion paper*, **20/2000**, of the Norwegian school of economics and business administration.

[4] Burger M., Klar B., Müller A. and Schindlmayr G. (2004): A spot market model for pricing derivatives in electricity markets. *Journal of Quantitative Finance*, **4**(1), 109-122.

[5] Deng S. (1998): Stochastic models of energy commodity prices and their applications: Mean reversion with jumps and spikes. *PSERC report 98-28*.

[6] Deng S., Johnson B. and Sogomonian A. (1998): Spark spread options and the valuation of electricity generation assets. *Working paper*.

[7] Elliott R. J., Sick G. A. and Stein M. (2000): Pricing electricity calls. *Preprint*.

[8] Eydeland A. and Geman H. (1998): Pricing power derivatives. *Risk magazine*.

[9] Hinz J. (2003): Modeling day-ahead electricity prices. *Applied mathematical finance*, **10**(2), 149-161.

[10] de Jong C. and Huisman R. (2002): Option formulas for mean-reverting power prices with spikes. *Discussion paper*, **251**, from Erasmus Research Institute of Management (ERIM), Erasmus University Rotterdam.

[11] Keppo J. (2004): Pricing of electricity swing options. *Working paper*.

[12] Kaas R., Goovaerts M., Dhaene J. and Denuit M. (2001): *Modern actuarial risk theory*. Kluwer publishing.

[13] Koekebakker S. and Ollmar F. (2001): Forward curve dynamics in the Nordic electricity markets. *Working paper*, **21**, from the Norwegian school of Economics and business administration.

[14] Longstaff F. and Wang A. (2002): Electricity forward prices: A high frequency analysis. *Finance pape* **10-02**, July 1.

[15] Lucia J., Schwartz E. (2000): Electricity prices and power derivatives: Evidence from the Nordic power exchange. *Finance paper* **16-00**, March 1.

[16] The Nordic power exchange (2003): `http://www.nordpool.com`.

[17] Verschuere M. (2003): `http://www.fam.tuwien.ac.at/˜michel`.

[18] Verschuere M. (2003): A comment on a regime switching model for power options, submitted.

[19] Schaden M., von Gravenstein L. and Verschuere M. (2003): On the anomalous behavior of power futures, submitted.

[20] Verschuere M. and von Grafenstein L. (2003): Futures hedging in power markets: Evidence from the EEX. *Working paper*.

[21] Stoft S., Belden T., Goldman C. and Pickle S. (1998): Primer on electricity futures and other derivatives. *Electricity market studies* **LBNL-41098**.

[22] Teichmann J. (2003): A Note on Non-Affine Solutions of Term Structure Equations with Applications to Power Exchanges, submitted.
`http://www.fam.tuwien.ac.at/˜jteichma`

[23] Weron R., Simonsen I. and Wilman P. (2003): Modeling highly volatile and seasonal markets: Evidence from the NordPool electricity market, *Econometrics* **0303007**, Economics working paper archive at WUSTL.

INTEGRATION OF A PRIORI SEGMENTATION IN BONUS-MALUSSYSTEMS

Ingrid Vreven

Generali Belgium, Louisalaan 149, 1050 Brussel
Email: ingrid.vreven@generali.be

Abstract

The old Belgian legal bonus-malussystem does not accomplish the two goals of a bonus-malussystem, which are: reaching a financial balance in the company and anticipation of the risk the insured brings to the company. This paper presents a model that does fulfil both goals by integrating a priori segmentation into the bonus-malussystem.

1. INTRODUCTION

To determine the premium of an insurance risk, the risk should first be evaluated. There are two important ways for evaluating a risk in a car portfolio: a priori, where we, by the use of criteria, such as the age and the sex of the insured, evaluate the risk before the insured is able to drive on the road. The goal of the a priori segmentation is to distinguish different homogeneous risk classes: the insured in the same risk class pay the same base premium. But even if we use many criteria it is impossible to make a correct prediction of the claim frequency only by the use of a priori classification, because the portfolio is heterogeneous: there are things in a car portfolio that can't be measured, such as the driving behaviour of the conductor. That is why there is also need of an a posteriori evaluation, where we take into account the claim history of the insured. The a posteriori evaluation is done by the use of a bonus-malus system. The goal of a bonus-malus system is the correction of an a priori wrong judged risk by an increment or a decrement of the premium.

A car insurance portfolio is heterogeneous. We are, by the use of criteria never able to predict how a certain risk will behave. The difference between the real number of accidents and the predicted number of accidents is called the heterogeneity. By a priori classification the heterogeneity will decrease, because the predictions are based on data of a select group instead of on data of the whole portfolio. But a bonus-malus system remains necessary, because the heterogeneity can't be eliminated.

A bonus-malussystem has two important goals for the insurance company:

1. a better anticipation of the risk, so that every body pays, after a while, a premium that corresponds to his own claim frequency, and

2. the keeping of a financial balance in the company.

Both goals are not accomplished in the old legal system: an insured with a claim frequency of 12% pays less than 20% more than an insured with a claim frequency of 10% and the average premium level of an insured with a claim frequency of 10% keeps decreasing, because the insured has reached the lowest level, so there can never be a financial balance in the company.

The reason why the old legal system does not accomplish those demands is that there is only one bonus-malussystem for the whole portfolio. This is actually wrong, because after a priori classification, the remaining heterogeneity is to be found at the level of the risk classes, so it is more appropriate to use a bonus-malusscheme for each risk class.

Based on this idea, Gisler developed a model where he combines the a priori classification and the a posteriori evaluation. The model of Gisler is based on the credibility theory, more specific the Buhlmann-Straub model, where the credibility predictor is given by a linear combination of the collective predictor and the individual predictor for each risk class.

2. MODEL WITHOUT A PRIORI SEGMENTATION

Define N_{ij} as the number of claims of risk i in year j. Each risk i has a risk parameter λ_i, the claim frequency. The average claim frequency of the whole portfolio is represented by λ. Define $\vartheta_i = \frac{\lambda_i}{\lambda}$, ϑ_i is the real bonus-malus level of risk i: it is the level in which the risk i is better or worse than the average risk.

In order to get a better representation of the results, we define $\widetilde{N}_{ij} = \frac{N_{ij}}{\lambda}$. Then $\widetilde{N}_{ij}$ fulfils the conditions of the Buhlmann-Straub model, so the credibility estimator for ϑ_i can be determined with it. The credibility weight is equal to:

$$z_i = \frac{n\lambda}{n\lambda + w^{-1}} \,,$$

with:

- weights equal to the number of years n,

- $\text{Var}[\mu(\vartheta_i)] = w$, the heterogeneity,

- $\text{E}[\sigma^2(\vartheta_i)] = \lambda^{-1}$.

Because in this model $m = \text{E}[\mu(\vartheta_i)] = 1$ and the individual estimator of the claim number is equal to $\frac{\sum_j \widetilde{N}_{ij}}{n}$, the credibility estimator is equal to

$$\begin{aligned}\hat{\vartheta}_i &= \left(1 - \frac{n\lambda}{n\lambda + w^{-1}}\right) + \frac{n\lambda}{n\lambda + w^{-1}} \frac{\sum_j \widetilde{N}_{ij}}{n} \\ &= 1 + \frac{n\lambda}{n\lambda + w^{-1}} \left(\frac{\sum_j \widetilde{N}_{ij}}{n} - 1\right). \end{aligned} \tag{1}$$

$n\lambda$ is the a priori expected claim number within the observation period of n years. Without more information we expect that an insured with claim frequency λ in a period of n years will cause $n\lambda$ accidents.

The equation (1) is also equal to:

$$\hat{\lambda}_i = \lambda + \frac{n}{n + (w\lambda)^{-1}} \left(\frac{\sum_j N_{ij}}{n} - \lambda\right).$$

The structural parameters λ and w can be determined from the data of the portfolio. When I is the number of contracts in the portfolio, than λ and w can be estimated as follows:

$$\begin{aligned}\hat{\lambda} &= \frac{1}{I} \sum_i N_i \\ \hat{w} &= \hat{\lambda}^{-2}(\hat{\sigma}^2_{N_i} - \hat{\lambda}),\end{aligned}$$

with $\hat{\sigma}^2_{N_i} = \frac{1}{I-1} \sum_i (N_i - \hat{\lambda})^2$ (no index j because we observe only one year).

This model assumes that all risks have the same a priori claim frequency and that differences in claim frequency of the risks are due to the individual risk characteristics ϑ_i. According to this model an insured of 20 years old, who is inexperienced, drives with a sport car and uses his car mostly in the city has as much probability of causing an accident then somebody who is 40 years old, experienced, drives with a family car and uses his car mostly at the country side. This is of course not very realistic.

3. MODEL WITH A PRIORI SEGMENTATION

The previous model doesn't take the differences between the risks due to the profile of the insured or the type of the car into account. We can adapt the previous model by assigning to each risk a parameter ϑ_i and a claim frequency λ_{ij}, depending on the year j. Instead of working with the global average claim frequency, we take the average claim frequency over the years j of each risk i into account.

Define N_{ij} as the number of claims of risk i in the year j, and as in the previous model: $\widetilde{N}_{ij} = \frac{N_{ij}}{\lambda_{ij}}$, then the conditions of the Buhlmann-Straub model are fulfilled and the credibility estimator is equal to:

$$\hat{\vartheta}_i = 1 + \frac{\lambda_{i\cdot}}{\lambda_{i\cdot} + w^{-1}} \left(\frac{N_{i\cdot}}{\lambda_{i\cdot}} - 1\right) \tag{2}$$

with $\lambda_{i\cdot} = \sum_{j=1}^{n} \lambda_{ij}$ and $N_{i\cdot} = \sum_{j=1}^{n} N_{ij}$.

Formula (1) and formula (2) look very similar, but there are two major differences:

1. Formula (2) takes the claim frequency of each risk separately into account. If we would a priori separate the risks in risk classes, we would find a bonus-malus system for each risk class. Instead of the claim frequency for each risk we take the average claim frequency for each risk class.

2. In formula (2) the a priori expected claim number is given by $\sum_j \lambda_j$. The formula (2) can change based on the a posteriori variables, which are taken into account in the price of the insurance. This was not possible in formula (1), where this number was given by $n\lambda$.

4. INFLUENCING FACTORS

The larger the expected claim number, $\lambda_{i\cdot}$, the larger the credibility weight. If for instance the expected claim number of a risk in class i is only half the number of a risk in class j, then the risk in class i needs two times more time to reach the same premium percentage as a risk in class j.

If $\lambda_{i\cdot}$ is very small, which means that there are very little accidents caused by the insured in that class, then the credibility weight will also be very small, and consequently the bonus of a risk without claims is also small. But if an accident happens the bonus-malus factor will increase drastically and bring a high malus. The reason is that an insured, in a class with a low claim frequency, already pay a much lower base premium than an insured in a class with a high claim frequency. If an accident happens the malus is that high because the insured maybe belongs in a class with a higher claim frequency with a higher base premium. The opposite is also true. An insured in a class with a higher claim frequency and so a high base premium, who drives for years without accident, gets a high bonus because he should be in a class with a lower claim frequency, where he pays a lower base premium.

Another factor that influences the bonus-malussystem in each risk class is the remaining heterogeneity. In case of a high heterogeneity the probability of a misjudgement is bigger than in the case of a low heterogeneity, so there is more need of a posteriori corrections in case of high heterogeneity.

5. NUMERICAL EXAMPLE

We separate the portfolio by the age of the insured, the year in which they got their licence and the zone in which they live. The a priori expected claim frequency is determined by the use of a generalised linear model (for details see Vreven [4]).

Age	Licence	Zone	Number of contracts	Claim frequency
18 - 22	1999 - 2001	country	733	0,121
18 - 22	1999 - 2001	city	561	0,158
23 - 29	1990 - 1998	country	3.297	0,08
23 - 29	1990 - 1998	city	2.618	0,104
23 - 29	1999 - 2001	country	724	0,145
23 - 29	1999 - 2001	city	510	0,173
30 - 69	before 1990	country	60.360	0,051
30 - 69	before 1990	city	21.512	0,067
30 - 69	1990 - 1998	country	4.416	0,071
30 - 69	1990 - 1998	city	3.331	0,105
30 - 69	1999 - 2001	country	78	0,156
30 - 69	1999 - 2001	city	59	0,182
70 - 103	before 1990	country	4.129	0,05
70 - 103	before 1990	city	3.095	0,06
70 - 103	1990 - 1998	country	24	0,161
70 - 103	1990 - 1998	city	18	0,176
70 - 103	1999 - 2001	country	2	0,194
70 - 103	1999 - 2001	city	1	0,225
Total			105.468	$\lambda = 0,062$

Table 1: Distribution portfolio

We multiply the a priori expected claim frequency with the average cost of a claim for the company. Assume that the average cost is equal to 4.155 Euro, than the base premium for an insured of 25 year old, who got his licence in 1995 and lives at the country side (zone 1) is 329,20 Euro. For an insured with the same age and licence year, who lives in the city (zone 2), the base premium is equal to 433,32. After 5 years the insured reach their 30 years, so they change risk class. Their base premium becomes 315,99 Euro in zone 1 and 415,92 Euro in zone 2.

Insured living zone 1							
t	BP$_t$	0 claims		1 claim		2 claims	
		BMF	Premiums	BMF	Premiums	BMF	Premiums
1	329,20	89%	292,99	228%	750,58	366%	1.204,87
2	329,20	80%	263,36	205%	674,86	330%	1.086,36
3	329,20	73%	240,32	186%	612,31	300%	987,60
4	329,20	67%	220,56	171%	562,93	275%	905,30
5	329,20	62%	204,10	158%	520,14	254%	836,17
6	315,99	60%	189,59	154%	486,25	248%	783,66
7	315,99	56%	176,95	144%	455,03	232%	733,10
8	315,99	53%	167,47	136%	429,75	218%	688,86
9	315,99	50%	158,00	128%	404,47	206%	650,94
10	315,99	47%	148,52	121%	382,35	195%	616,18

Table 2: Bonus-malusfactors and a posteriori premiums for a 25 year old conductor living in zone 1

The first column represents the number of years t, the second column (BP_t) contains the base premium, the third column (BMF) represents the bonus-malusfactor in case the insured causes no accident in the period $[0, t]$. The fourth column gives the pure premium. The next two columns represent the bonus-malusfactor and pure premium for an insured who causes one accident during this period and the last two columns give the bonus-malusfactor and pure premium for an insured who causes two accidents during this period.
The a posteriori premiums are the product of a base premium, depending on the personal characteristics of the insured and a bonus-maluscoefficient. This bonus-maluscoefficient is also depending on the personal characteristics of the insured. A 25 year old driver, who got his licence in 1995 and lives in the city, has another base premium, but also other bonus-malusfactors than the 25 year old driver, who got his licence in 1995, but lives at the country side.

Insured living zone 2							
t	BP_t	0 claims		1 claim		2 claims	
		BMF	Premiums	BMF	Premiums	BMF	Premiums
1	433,32	86%	372,83	220%	954,45	354%	1.537,07
2	433,32	76%	327,16	193%	837,54	311%	1.347,91
3	433,32	67%	291,46	172%	746,14	277%	1.200,82
4	433,32	61%	262,78	155%	672,73	250%	1.082,67
5	433,32	55%	239,24	141%	612,47	227%	985,69
6	415,92	52%	216,28	132%	549,01	213%	885,91
7	415,92	48%	199,64	122%	507,42	197%	819,36
8	415,92	44%	183,00	114%	474,15	183%	761,13
9	415,92	42%	174,69	106%	440,87	171%	711,22
10	415,92	39%	162,21	100%	415,92	161%	669,63

Table 3: Bonus-malusfactors and a posteriori premiums for a 25 year old conductor living in zone 2

When we compare both tables, we note that the 25 year old driver living in zone 1 will always have a lower base premium than the 25 year old driver living in zone 2. But the insured from zone 1 has higher bonus-malusfactors, so he gets fewer bonuses and more malus than the insured from zone 2. This is a consequence of the fact that good risks get already a bonus by paying a lower base premium, so the level of his real bonus decreases. Or, the insured who are a priori wrong classified in a certain risk class, because their claim frequency is lower than the claim frequency of the other risks, but more comparable with the claim frequency of the insured in another class, get a bonus that is large enough that after a certain number of years they pay as much as the insured from the right risk class.

6. EVALUATION OF THE SEGMENTATION MODEL

As already mentioned, a bonus-malussystem has two important goals for the insurance company: a better judgement of the risks, that everybody pays, after a certain time, a premium that corresponds with his own claim frequency and the keeping of a financial balance in the company. The level in which these goals are accomplished is different in each bonus-malussystem. That's why we compare the old legal system and the segmentation model by the methods described by Lemaire: the average premium level and the variation of the premium. Also the number of assigned bonuses and maluses in both systems were counted.

The comparisons are made by the use of simulation programs in SAS. These simulation programs use a database in which every insured has his claim frequency l. To simulate the old legal system, we let everybody start in class eleven. After one year, the insured had zero, one or two or three accidents, simulated with a random number from a Poisson distribution. It is also possible that the insured leaves the company for competitional reasons or that the insured dies with the consequence that the policy disappears from the portfolio. Also new customers enter in the portfolio. If we assume that the number of policies that disappears from the portfolio is equal to the number of policies that enters the portfolio, we can simply solve this by assigning a random number between zero and one, simulated with a uniform$(0, 1)$ distribution. When this number is smaller than a certain threshold, depending from company to company, we assume that this policy leaves the portfolio and enters in bonus-malus eleven. The threshold was determined by comparing the distribution of the portfolio over the different bonus-maluslevels obtained by the simulation with the real distribution of the portfolio.

An important property of the old legal system is the high concentration of insured in the lowest classes, the classes with the highest discounts. The first simulation program calculates for the old legal system and for the system with segmentation, the average premium level after one hundred years of an insured with a claim frequency of 10%. The result is shown in figure 1.

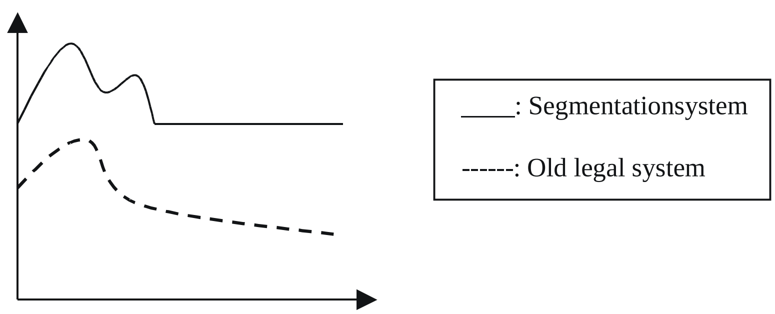


Figure 1: Average premium level in case of a claim frequency of 10% over 100 years

The goal of reaching a financial balance is not accomplished in the old legal system: the average premium level keeps decreasing. On the contrary in the system with segmentation we reach a financial balance after 20 years. After this time the insured pays a premium that corresponds with his own claim frequency.

The second comparison we’ve made is the number of bonuses and maluses in each system. Because most of the people are in the lowest classes in the old legal system, the number of bonuses assigned by the company is much higher than the number of maluses, which causes also a disturbance in the financial balance of the company.

The second simulation program counts the number of insured in the portfolio that pays after 30 years less than 100% of the premium and the number of insured that pays after 30 years more than 100% of the premium and this for the old legal system and for the system with a priori segmentation.

	Number	Average Level
BONUS: premium $<$ 100%	100.298	0,59
MALUS: premium $>$ 100%	3.718	1,42

Table 4: Bonuses and maluses in the old legal system

	Number	Average Level
BONUS: premium $<$ 100%	66.242	0,64
MALUS: premium $>$ 100%	37.774	1,63

Table 5: Bonuses and maluses in the segmentation system

We can see that the old legal system assigns at portfolio level more bonuses than maluses. Also, with the segmentation system, the number of bonuses is larger than the number of maluses, but the difference between both is much smaller. The average bonus level is also smaller in the system with segmentation, while the average maluslevel is higher. This fact together with the information of the average premium level teaches us that the old legal system does not accomplish the goal of a financial balance, in contradiction to the system with segmentation, where, after a while a financial balance is obtained.

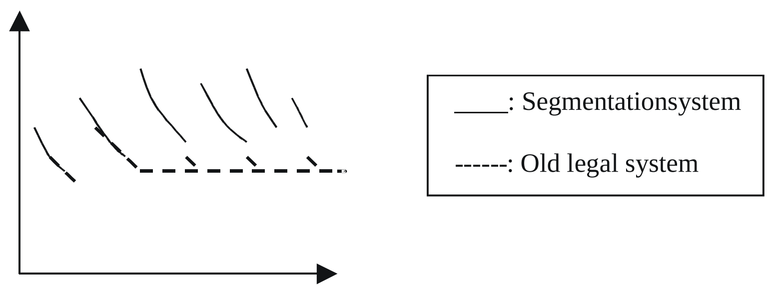


Figure 2: Variation on the premium

By the a posteriori corrections on the premium, the payments of the insured will be different from year to year according to their claim history. The variation of the premium was simulated for both systems. Figure 2 gives us the variation on the premium, an insured with a claim frequency of 10% would pay, during 60 years. We see that the variation of the premium in the system with segmentation is much bigger than in the old legal system, where the insured has reached the lowest class and causes now and then an accident. The variation in the system with segmentation is high, although there is a slight decrease noticeable.

7. CONCLUSION

To determine the premium of an insurance risk, the risk should first be evaluated. When the risk is wrong evaluated, it can have consequences for the insured but also for the insurance company. That is why it is important that the insured pays a premium that corresponds exactly with the risk he brings for the company. With the a priori segmentation we can already judge the risk from the subscription. By the a posteriori corrections the insured pays, after a while, a premium corresponding with his own claim frequency, expressed by the average premium level that is equal to 100%.

By comparing the number of bonuses and maluses in the old legal system and in the segmentation model, we notice that the old legal system assigns a lot of bonuses and little maluses. Although the goal of a bonus-malussystem is the distinction of the good and the bad risks. We can conclude that the old legal system is less effective than the model with segmentation. The old legal system distincts only the real bad and the less bad risks, but we cannot detect the good risks. In the model with segmentation on the contrary, the malus is that high that we know immediately which ones are the good risks and which ones the bad.

Despite the being more correct of the segmentation system, there are a few disadvantages on the segmentation model. There is a lot of variation on the premium, so the insured needs a lot of time to reach his old premium level again after an accident.

References

[1] Gisler A. (1996): Bonus-malus and tariff segmentation. Paper presented at the 27th ASTIN Colloquium, Copenhagen.

[2] Goovaerts M.J, Kaas R., Van Heerwaarden A.E (1990): *Effective actuarial methods*. North-Holland, Amsterdam, 316 p.

[3] Lemaire J. (1999): Bonus-malussystems: the European and Asian approach to merit-rating. *North American Actuarial Journal*, **2**, 26-44.

[4] Vreven I. (2002): Integratie van a priori segmentatie in bonus-malussystemen. *Thesis actuariële wetenschappen*, VUB.

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 "2nd Actuarial and Financial Mathematics Day" (6 februari 2004, Prof. M. Vanmaele)

De "2nd Actuarial and Financial Mathematics Day" was net als de vorige editie 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. Daarnaast biedt het contactforum een mogelijkheid om de resultaten van het uitgevoerde onderzoek aan de praktijkmensen uit banken en verzekeringen – die in ruime getale aanwezig waren – voor te stellen. Naast twee uitgenodigde sprekers kwamen doctoraatsstudenten, postdocs evenals een spreker uit de praktijk aan het woord.
In deze publicatie vindt u een neerslag van de voorgestelde onderwerpen zoals het prijzen van samengestelde opties en Aziatische opties, interestmarktmodellen, afgeleide producten in de energiemarkt, benadering van de distributie van annuïteiten in het geval van stochastische rentevoeten, evenals een benadering voor het probleem van een optimale portefeuilleselectie, analyse van het risico van kredietportefeuilles, een aanpak voor gecorreleerde risico's in actuariële problemen en segmentatie in bonus-malussystemen.