KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

4TH ACTUARIAL AND FINANCIAL MATHEMATICS DAY

February 10, 2006

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

CONTACTFORUM

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

4TH ACTUARIAL AND FINANCIAL MATHEMATICS DAY

February 10, 2006

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

CONTACTFORUM

Handelingen van het contactforum "4th Actuarial and Financial Mathematics Day" (10 februari 2006, 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/2006/0455/20

Printed by *Universa Press, 9230 Wetteren, Belgium*

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

4th Actuarial and Financial Mathematics Day

PREFACE

The Contactforum “Actuarial and Financial Mathematics Day” was organized for the fourth time. It started four years ago with a modest meeting but since then, this event attracted every year more and more participants.
The main purposes of this event is twofold. Firstly, we want to bring people together from two fields with a lot in common, namely the actuarial field and the financial field. This is important seen the recent evolution on a company level but also by looking at the nowadays battery of interrelated products such as equity-linked insurances and credit risk. Secondly, our aim is to bring practitioners and academics closer together in order to create a stimulating interaction for both of them. This edition welcomed as many practitioners as academics.
This contactforum gives on one hand young and promising researchers the opportunity to present their recent work to a broad audience and to have their paper published in these proceedings. On the other hand, renowned practitioners were programmed as main speakers in order to give them a forum to talk about the needs, the problems, the hot topics in their fields. The invited paper about Solvency II is included in these transactions.

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

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

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

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

4th Actuarial and Financial Mathematics Day

CONTENTS

Invited talk

Contributed talks

INVITED TALK

SOLVENCY II :
AS SIMPLE AS POSSIBLE, AS COMPLEX AS NECESSARY
(THE STORY OF A PASSIONATE CHALLENGE FOR ACTUARIES)

René Dhondt

Assuralia, De Meeûssquare 29, 1000 Brussels, Belgium & Universiteit Antwerpen, Faculty of Applied Economics, Prinsstraat 13, 2000 Antwerp, Belgium
Email: `rene.dhondt@assuralia.be`

Abstract

The need to redefine prudential standards according to the real risks insurance companies are exposed to, has proven to be much more than pure intellectual work. Indeed, trading off between policyholders protection and optimal capital allocation is no easy task (neither technical nor political). The European Commission has to manage quite a complex project, taking into account many divergent point of views: 25 national supervisors of as many different domestic insurance markets (trying to coordinate within CEIOPS), representatives of insurance undertakings (also coordinating their points of views within CEA, AISAM, CFO/CRO-Forums, ...), consumers, accountants, auditors and ... last but not least the actuaries whose advice is expected from the "Groupe Consultatif Actuariel Européen". The Swiss Solvency Test could prefigure some major parts of the output of the Solvency II project, one of them concerning the "Standard Models" that CEA means to be "As simple as possible, as complex as necessary".

1. INTRODUCTION

In order to operate, an insurance company requires a number of different 'resources' (the description given here is simplified and limited in scope):

- An approach to the risks it has to deal with (not only insurance-related risks, but also various business and financial risks).
- An extensive historical knowledge of the frequency and magnitude of these risks, as they actually materialise.
- Staff with the necessary knowledge and experience (to undertake the various sub-activities-product development, marketing, distribution, production and loss management, legal and insurance-related support, management of all the above, etc.).

- The necessary financial resources - the only raw material that insurers use. These financial resources may be provided by shareholders, obtained on the financial or reinsurance markets or derived from the company's annual business activities.

Available capital is far and away the most important element for an insurance company. Capital is a basic commodity for insurers, allowing them to accept risks. It must be used effectively and in an optimum way. If the necessary capital has to be borrowed or bought, it has a price, as have all goods handled on a supply and demand market. This price, as well as the capacity of the financial markets, are not defined at national level, not even at European level, but globally.

Identifying the precise capital needs of an insurance company is a complex exercise. The main reason for this is that the normal production cycle is reversed: in other words the real cost of the insurance product is only calculated when the risk materialises, not when the insurance contract is drawn up. Indeed, it is highly likely that the price (or 'premium') asked for an insurance contract will differ greatly from the real (average) cost of the insurance guarantee.

European legislators, realising from the outset that a minimum degree of harmonisation was necessary to the development of the single market, developed an initial set of rules known as Solvency I. Solvency I, which came into being at the time of the 'first generation' directives and still officially applied by national supervisors, is purely quantitative in nature; solvency requirements are expressed as a fixed percentage of the earned premiums and/or the loss provisions. Naturally, this makes it impossible to properly study the solvency risk of different insurance companies and how this risk is influenced by, say, the legal form of the company, its investment policy and its product definition and price setting mechanisms. Further, the Solvency I rules provide for inadequate harmonisation, so that insurers operating from different Member States cannot be considered in the same framework of reference. Finally, these rules were drawn up in the 1970s and so obviously take no account of new concepts such as asset and liability management, alternative risk transfer or the existence of derivatives offering financial protection.

In view of this, a new project, Solvency II, was launched by the European Commission in the late 1990s. This project is about providing the capital needed to guarantee the continuity of insurance companies and giving near-certainty to beneficiaries that the payments will be executed in due time.

Calculating the Solvency II margin for the insurance industry is one of the greatest adventures in the financial world at the start of the third millennium. It is a voyage of discovery for commodities, as important as those undertaken in the late 15th and early 16th century.

2. THE SOLVENCY OF THE BELGIAN INSURANCE MARKET

Let us now examine the solvency of the Belgian insurance market in real terms. What can be legitimately expected from a solvency model?

First of all, the Solvency I rules, as applied in Belgium, seem to be effective, since there has been no bankruptcy in the Belgian market in the last 30 years (the Insurance Companies Supervision Act, drawn up in response to the first-generation directives, dates from July 1975). This cannot be pure chance, since a number of companies in neighbouring countries have gone under in the recent past: the Belgian market clearly has a financially solid foundation.

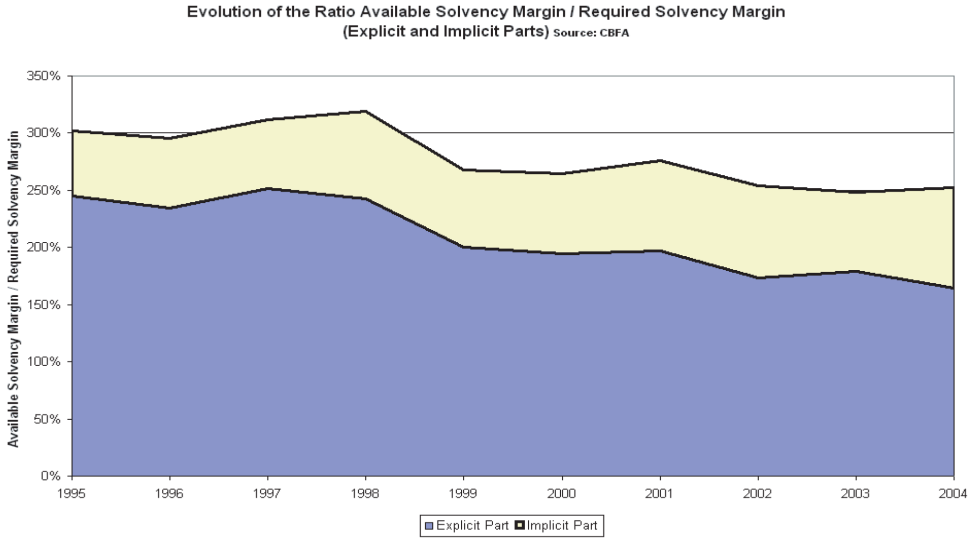


Figure 1: Evolution of the Ratio Available Solvency Margin / Required Solvency Margin

Let's take a more detailed look at the Belgian market since 1995.

Figure 1 illustrates, over the period from 1995 to 2004, the ratio between the real overall available solvency margin of the Belgian direct insurance market (not including reinsurance) and the margin imposed by the Solvency I rules. The Belgian market has seen its solvency margin decline slightly from 300% to 250%, but it remains comfortably outside the danger zone. The solid 'explicit' part of this margin (i.e. the part based on the market value of underlying assets) has fallen slightly more markedly than the more volatile 'implicit' part (unrealised gains, expected profits, etc.).

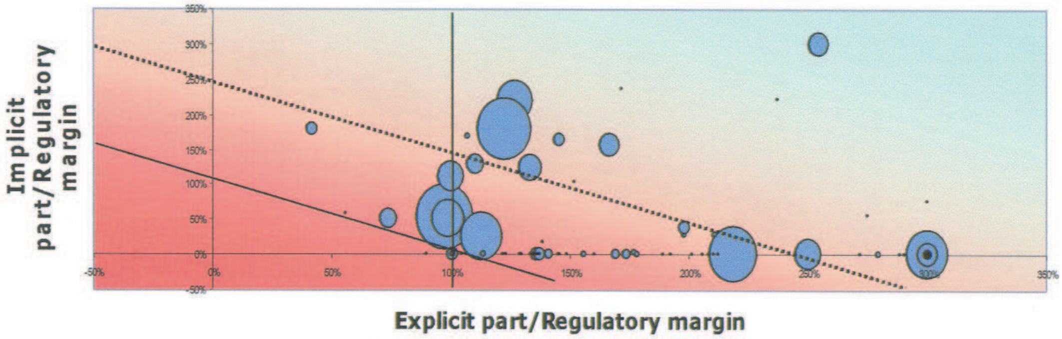


Figure 2: Dispersion of Solvency Margins

This positive overall situation gives us no definite indication about the solvency of individual companies. For clarification on this, we need to look at the spread of individual margins (figure 2). The horizontal axis shows the explicit part (as a percentage of the regulatory margin, with maximum value 300%), the vertical axis shows the implicit part and each circle represents one insurance company. The line passing through the 100% level on the two axes represents the regulatory minimum. No company is beneath this level, which means that every company possesses at least the minimum required solvency capital. The dotted line shows the average margin actually available. The companies are spread out around this average value but it is interesting to note that several companies have a solvency capital, or equity, over three times higher than the required capital level.

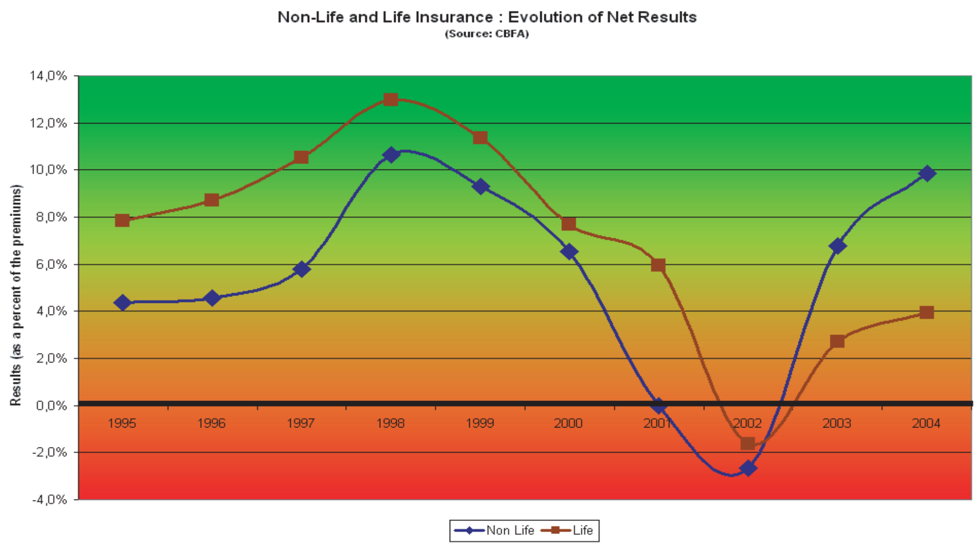


Figure 3: Evolution of Net Results

The increase in solvency capital is financed by annual financial revenues from underlying insurance transactions. Figure 3 shows how, following a real increase in companies' net results which lasted until 1998, there ensued a period of decline: initially, until 2000, results remained very positive, but then came two successive years of loss in non-life insurance and – for the first time ever – a year of overall loss for life-insurance companies (2002).

Accounting results remained positive until the turn of the Millennium, but the continual destruction of economic value in the insurance industry soon became apparent.

In 2000, the industry's professional organisation joined forces with McKinsey & Company to lead an awareness-raising campaign on 'economic capital'. Some companies were already very aware of the concept and had firmly embedded it in their business policies; for others, it was still relatively unfamiliar.

Few of them had already integrated a model centred around the 'embedded value' of the portfolio, and there was too little awareness of the danger of underestimating the risk of value destruction. Strong competition (inadequate pricing in non-life business, too high interest rates guaranteed in life business) led to a decrease in profits and later, as a consequence of the tumbling stock markets (from 2000), gave negative results, thus affecting the available capital and the solvency margins.

A look at the combined equity of the Belgian insurance industry between 1998 and 2004 (figure 4) shows that the book value remained roughly the same but the losses sustained by insurers in 2001 and 2002 had major repercussions on this value, which led to capital injections for many companies in 2003 (in total between €550 and 600 million). The collapsing capital markets clearly had a destructive impact on companies' equity, but other factors too contributed to the loss, in the space

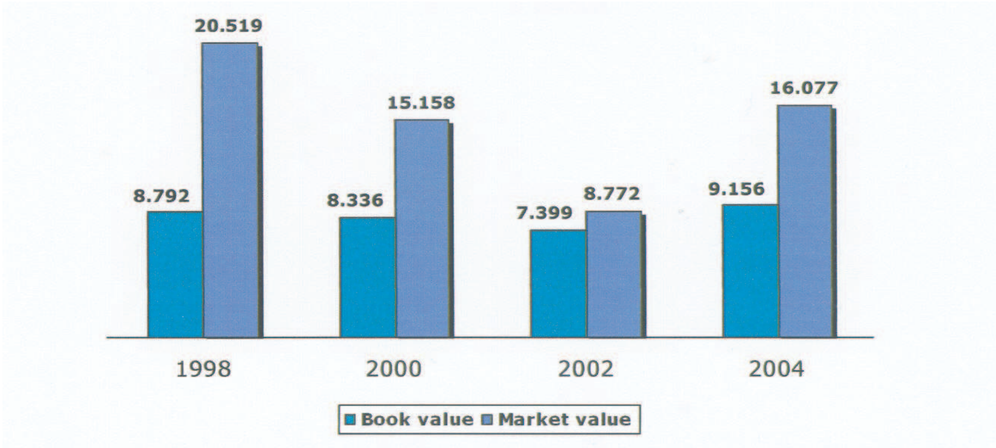


Figure 4: Evolution of Equity Capital (in mio euro)

of four years, of over a half of the previously amassed equity, measured in market value terms, from €20.5 to 8.8 billion. The trend was broken in 2003 when significant positive results were again recorded, leading to the recovery of over half of the lost equity. The sector would therefore seem to be resilient enough to ensure its own financial continuity. However, the volatility that has been displayed raises questions about the capital requirements imposed on companies and about the imminent reform of the financial reporting system, both of which could amplify this effect.

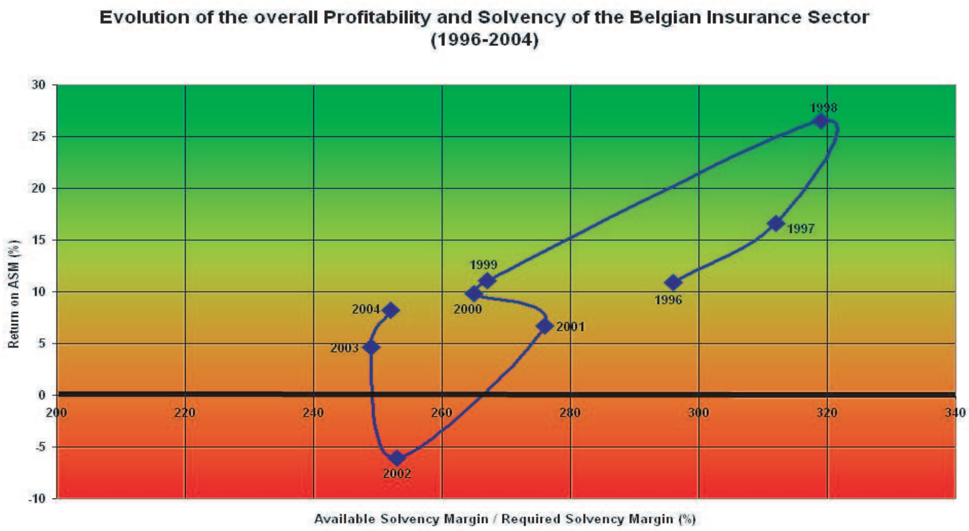


Figure 5: Evolution of Profitability and Solvency

Figure 5 summarises how the profitability and solvency of Belgian insurance companies have changed over the past nine years. The year 1998 stands out as a record year for the Belgian insurance industry, as regards the ratio between the available solvency margin and the required solvency margin and as regards the return on equity (return on available solvency margin). The years from 1999 onwards saw a sharp decline in profitability, with a negative return in 2002. This meant that the established solvency margin fell markedly in 2002 and 2003, without however approaching the required minimum.

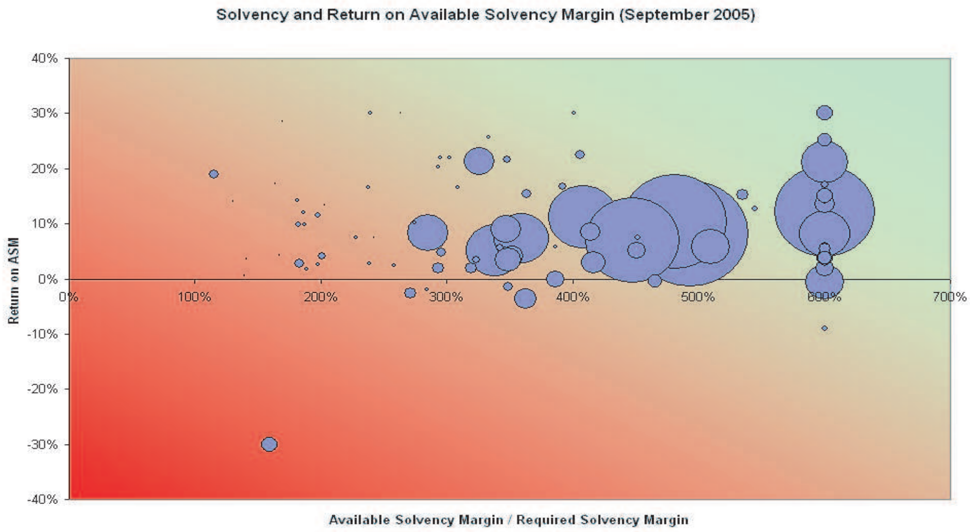


Figure 6: Solvency and Return on Available Solvency Margin

A look at the market situation gives an indication, although not a definite one, of the situation of individual companies. In figure 6, the same data are split up to show the position of individual companies. As before, each circle represents one insurance company. The ratio between the available margin and the required solvency margin (shown on the X-axis) is capped at 600%, while the return on the available margin (Y-axis) goes no higher than 30%.

Like figure 2, this diagram shows that all insurance companies possess a solvency margin deemed to be adequate (i.e. greater than 100% of the required margin). The return on this margin – a reliable indicator of a company's equity – is positive for most companies, but varies greatly from one to another.

In some situations, there may be a degree of doubt as to the adequate allocation of capital or the profitability of the underlying business. Or both. On the one hand, a high solvency margin might lead to a very positive quotation by the rating agencies (Standard & Poor's, etc.) and will give access to interesting contracts, especially for the insurance of industrial risks. On the other hand, high capital requirements may frighten stakeholders and lead them to prefer investments in other services, industry areas, or other commodities.

3. WHAT ARE THE (BELGIAN) INSURANCE INDUSTRY'S NEEDS WITH RESPECT TO SOLVENCY?

Belgian insurance companies possess sufficient equity to financially offset major discrepancies between the actual loss frequency and the theoretical expected value. Indeed, they have a comfortable surplus, measured in terms of the required solvency margin imposed by the Solvency I rules. This surplus is present at market level, but even individual companies seem in no immediate danger of encountering problems.

The existing solvency requirements are principally concerned with insurance-related divergences from the expected values resulting from statistical and actuarial calculations based on historic claims. Other ‘accidents’ resulting from, say, disappointing financial income (downturn on the share markets, long-term interest) or a prolonged interruption to activity, are not taken into account in the scenarios that managers or supervisors use to measure the business’s risk sensitivity so as to augment capital requirements accordingly. As indicated above, the absence of such calculation methods often means that insurance companies become overcapitalised, which in turn damages shareholders’ interests and may induce them to disinvest. Furthermore, these solvency requirements are not harmonised throughout the EU, which can impact on competition between insurers from different Member Sates.

Take, for example, the recent demand made by some EU Member States to create ‘guarantee schemes’. The necessity of such schemes depends heavily on the prudent evaluation rules these Member States initiated: much of the financial protection will be provided by correctly calculated liabilities (reserves), a surplus of covering assets and an amount of unrealised gains; however, the home country’s legislation demands that each business line should be profitable (this is the case in Belgium, but may not be in some other Member States). This provides additional protection vis-à-vis foreign companies, which are only required to have ‘total account equilibrium’. Another question one might raise is whether the new solvency requirements will take into account existing guarantee schemes for determining the capital required.

In view of this, there is everything to be said for encouraging much greater harmonisation of supervisory rules, thus creating a level playing field between insurance companies competing on the single European insurance market.

This means that new calculation methods for solvency requirements are necessary. These methods must take account of increasingly diverse ruin scenarios and also factor in, amongst other things, the management itself and the soundness of the working and production methods instigated by the insurer. The time is therefore ripe for a new set of rules, Solvency II. There is now a strong need for robust models to define suitable solvency margins – models based on in-depth financial, statistical, actuarial and economic research.

Likewise, there is a strong need for a solid framework, guaranteeing continuity for these models: after all, insurers cannot alter their strategic decisions every few years.

4. HOW TO MOVE FROM SOLVENCY I TO SOLVENCY II?

As stated above, the Solvency II project was launched by the European Commission in the late 1990s. It was designed to assess the financial solidity of insurers on a more prospective basis, incorporating all the risks an insurance company may face and proposing a harmonised system at European level.

The objectives were ambitious but reasonable: to provide greater protection of policyholders’ and claimants’ interests whilst boosting economic growth by optimising the allocation of risk capital in the financial sector.

Various stakeholders, such as the Comité Européen des Assurances (CEA), have launched studies and projects to help ensure that the insurance industry takes an active role in developing the new rules. Before starting the actual discussions, some clarification work was clearly necessary to try to reach a common understanding and consensus on some of the major issues related to Solvency II. This ongoing project, launched in early 2005 but continuing this year and possibly also next year, will help the insurance industry to develop common views on objectives, the way to meet them and the solvency models that will help insurance companies to deal with the various risks they face.

The insurance industry has shown strong support for a Solvency II framework with the following aims:

- Enable an institution to absorb significant unforeseen losses and offer reasonable insurance to policyholders (Framework for Consultation on Solvency II).
- Contribute to a "better managed and more competitive insurance industry that can better perform its key function of accepting and spreading risk" (Commissioner McCreevy).
- Encourage a single European market for financial services.

The industry has also formulated some general principles which should be taken into consideration:

1. Insurers should be able to measure the risk to which they are exposed – which has repercussions on the requirement of risk-based capital – and take into account the insurance risks they have underwritten.
2. There should be maximum harmonisation across the European insurance markets. This means that individual Member States or local supervisory authorities would not be able to develop requirements that are more stringent than those defined at a European level.
3. The current solvency capital requirement should be replaced by a twofold requirement. Firstly, in relation to minimum capital, i.e. the level below which insurance activity should be put in run-off and, secondly, with regard to solvency capital (sometimes referred to as target capital), which defines the capital needs for an 'ongoing business' and below which intervention may be required from the supervisory authority. This would not necessarily result in new equity being injected, but would encourage the insurance company's managers to review and monitor the company's procedures and methods.
4. The new approach to calculating the solvency requirement will make existing hidden reserves (the result of an occasionally over-cautious approach) more transparent when determining provisions, unrealised capital gains etc.

In the meantime, the re-working of the existing rules on supervision of insurers' solvency should take into account the lack of harmonisation between countries and insurance markets, the rigidity of the current rules, which focus on a merely quantitative approach and penalise companies with a prudent provisions policy, and, finally, the poor integration of risks and opportunities associated with technical and financial innovations (such as ALM, ART and derivatives).

The first phase of the Solvency II project led to an integration of the Basel II requirements into the insurance environment, adopting but also adapting them, with a series of regulations for:

1. financial resources;
2. the prudential supervision process;
3. market discipline;

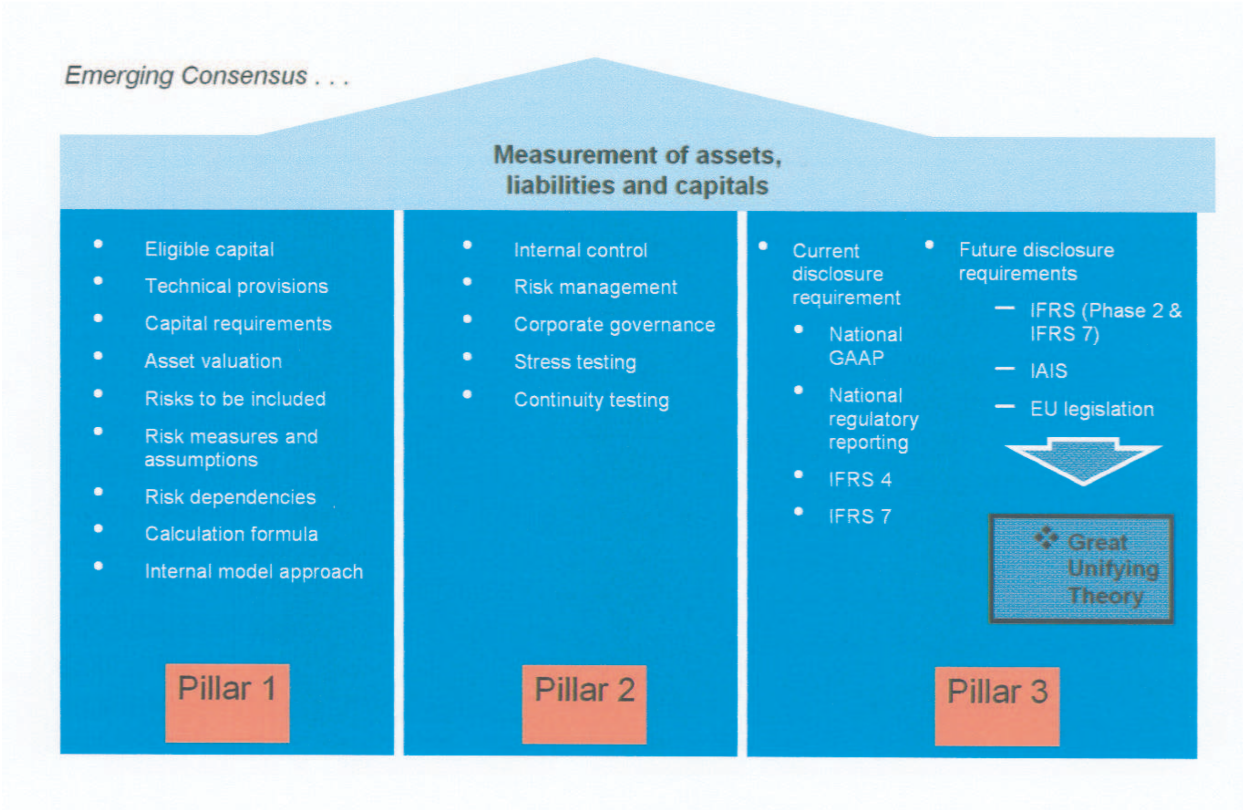


Figure 7: What Does the Solvency II Three-Pillar Approach Mean?

thus creating a three-pillar approach. The Pillar I Solvency capital requirements should be based on a total balance sheet approach, reflecting market-consistent value. This means measuring assets, capital and liabilities on a market value (if available otherwise they should be measured on the best estimate for projected future cash flow), see Figure 7. A sufficient level of harmonisation is necessary to ensure that solvency requirements are determined by the nature and scale of the activity and its risks rather than the location (or structure) of the company. Solvency rules should be designed for both financial groups and stand-alone companies and also require a lead supervisor to be appointed for the group.

A 'total balance sheet' approach involves evaluating all kinds of risks, taking into consideration not only statistical risk (probability of ruin due to unforeseen frequency or intensity of insured risks) but also financial risks, the risk of malfunctioning (improper product development, fraud etc.) and so forth. Each risk should be evaluated in a prospective way, on the basis of a stochastic evaluation of all future incoming premiums and payments.

The market value of the liabilities should be determined with enough certainty to ensure that no additional prudence is required to cover the risk of variation against the current market value. Although International Financial Reporting Standards (IFRS) have been incorporated into the accounting principles and rules used by insurance companies, and risk and capital management disclosures are part of this financial reporting system, IFRS and Solvency II clearly continue to differ considerably (IFRS equity versus regulatory capital as a result of the treatment of available-for-sale investments, scope of the group and its levels of consolidation as a result of the different understanding of the nature of insurance activities between banking and insurance). Accounting considerations must not affect the definitions used for solvency calculations.

The valuation of assets and liabilities and solvency capital requirement can be based on either an internal model, accepted by the supervisor, or a standard industry model. The standard approach will need to be more approximate and closer to the existing formulas. Insurers who don't have enough staff to develop their own internal solvency models should find the approach easy to apply. Moreover, it should also incorporate the recognition of diversification and risk mitigation.

Both methods should lead to the definition of two levels of equity capital requirements:

- A minimum capital requirement (MCR), calculated using simple formulas. This might even be the actual solvency margin and will also serve as an advance alert indicator.
- A target or solvency capital requirement (SCR) to reduce the probability that the company is bankrupted by a predefined level (e.g. 0.5% for a one-year period).

Internal models would lead to a reduction in capital requirements as regards the standard models and insurance companies or groups would only be allowed to use them under strict conditions (various aspects would be monitored, for example, the development and the quality of the models, the way in which they are used by the management, audit trails, the procedures followed and the level of compliance).

The standard model (a common insurance industry model) should be based on the same economic principles as an internal model but be simplified as far as possible. Each of the companies, even those using the standard model, should be encouraged to improve their risk management capacity. The practical limitations of the current 'one-size-fits-all' solvency requirement or of the Solvency II requirements could be avoided by using an approach that is more clearly based on principles rather than a traditional rules-based approach, e.g. with respect to investment rules (no arbitrary restrictions on investment flexibility needed since market, credit and liquidity risks are taken into account).

The total balance sheet approach, which incorporates results from either a standard or an internal model, is outlined in Figures 8 and 9. This still leaves a wide range of options available, as discussed below and in the definitions of the terms used in the figures.

Technical provisions are composed of:

- the actuarial best estimate: discounted provisions, without incorporated margin, decreased (for life provisions) with deferred acquisition costs and zillmerized, taking into account all options and guarantees;
- the market value margin, based on either:
  - the International Accounting Standards Board (IASB)'s tentative definition or;
  - the cost of transfer to a third party;
- a prudence margin, which might be either:
  - a specific percentage (function of the class of business) of the best estimate or;
  - a multiple of the standard error or;
  - a percentile of the probability distribution of the final costs (60/75 or 90%).

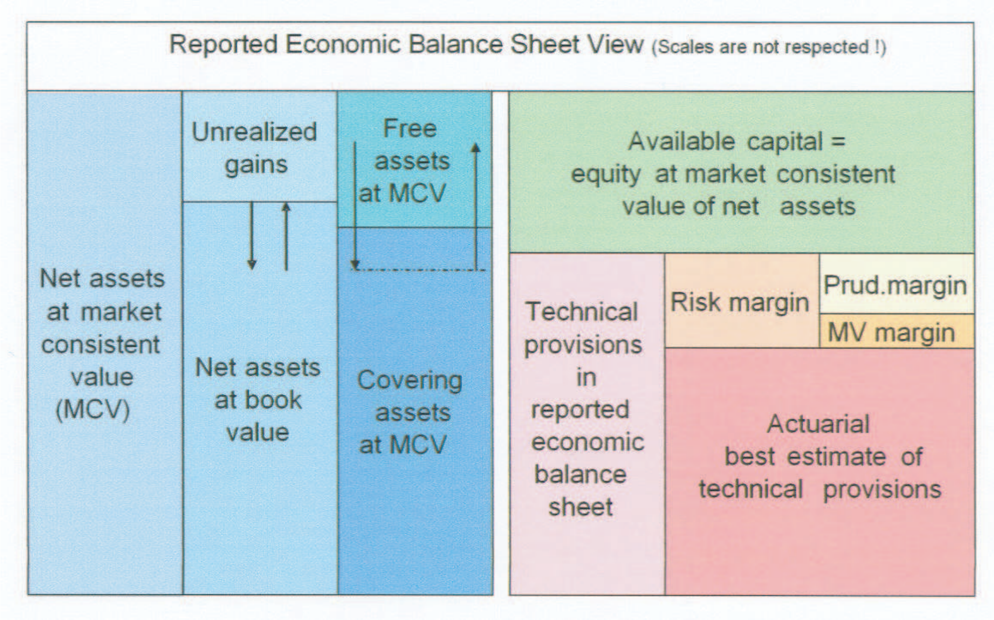


Figure 8: Total balance sheet approach (a)

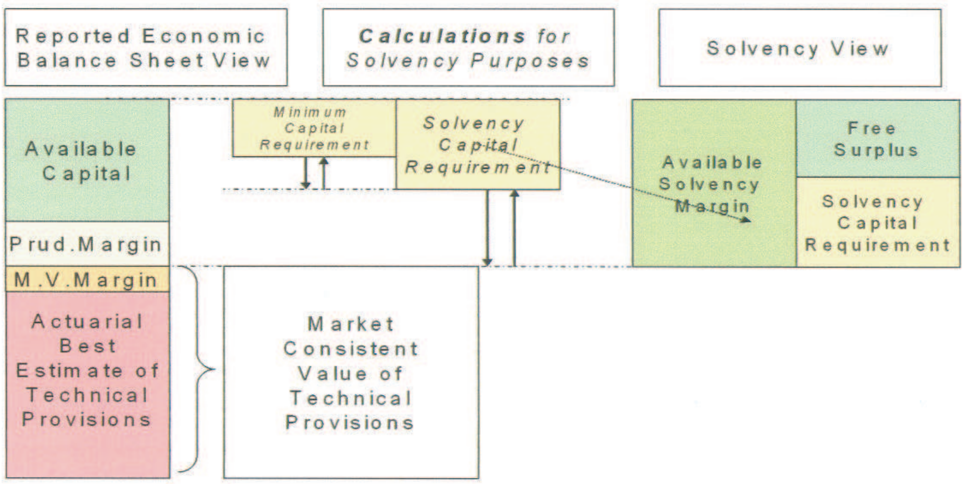


Figure 9: Total balance sheet approach (b)

Available capital (at market consistent value of net assets) is calculated by subtracting technical provisions from net assets:

$$\text{available capital} = \text{net assets} - \text{technical provisions}$$

with

$$\text{net assets} = \text{assets} - \text{debts},$$

where:

- assets and debts are calculated at market consistent value, which is the market value or marked to model (IASB's fair value?);

- intangibles are taken into account if they are not deducted from the liability side (goodwill, present value of future profits, net deferred taxes etc.).

Covering assets are reported either within the balance sheet or separately (separate annex to the supervisor?) and should always be superior to technical provisions and, if necessary, subject to harmonised (qualitative only?) investment rules.

Furthermore

$$\text{free assets} = \text{net assets} - \text{covering assets},$$

where a distinction could be made between assets representing MCR, SCR, free surplus and prudent-person principles could be adopted in relation to investment of some of these assets.

Measuring the economic value of liabilities requires calculation of the market-consistent value. This assumes that either market-consistent standards could be defined or robust internal models should be developed.

This process is not an easy one.

Almost all national insurance industries have started to develop market standards: new models that aim to issue an adequate total appraisal of the balance sheet appear every two months (e.g. Swiss Solvency Test, GDV models, FSA models, FFSA model to be launched or the Belgian model for workman's compensation insurance which is currently being developed).

Certain 'rules of thumb' could also be used, for example, a 75% confidence level for the best estimate of the technical provisions. (This means that there is 75% probability that the technical provisions would at least equal the pay-off for all liabilities in the portfolio).

Another possibility could be a predefined percentile (75%). It is for this reason that the Quantitative Impact Study (part 1 - QIS1), launched by national insurance supervisors on behalf of the CEIOPS (Committee of European Insurance and Occupational Pensions Supervisors), includes testing for the plus and minus 15% percentile, in order to gain a thorough understanding of the relation between the confidence level and the volume of technical provisions.

A predefined percentile does not give the competitive advantage that some insurance companies or groups aimed for when they set the development of an internal model as one of their key priorities. Other companies might not have of the necessary volume of technical data available to enter into the model.

Consequently, insurance associations like Assuralia are making preparations to help the market collect and collate this technical data.

As for matching assets with technical provisions, contributions will obviously be made by the International Accounting Standards Board (IASB) which defined the IFRS standards. So far, the insurance industry has been critical of the IFRS-standards which offered a disparate approach to liabilities and assets. The new IFRS rules (IFRS4 and IAS39) should allow a more consistent view on both sides of the balance sheet.

Figure 9 shows the possible outcome of the process as regards the level of the technical provisions (which will probably be lowered because the prudence margins would disappear) and solvency capital (where SCR would probably be less than the actual available solvency margin which would lead to a free surplus).

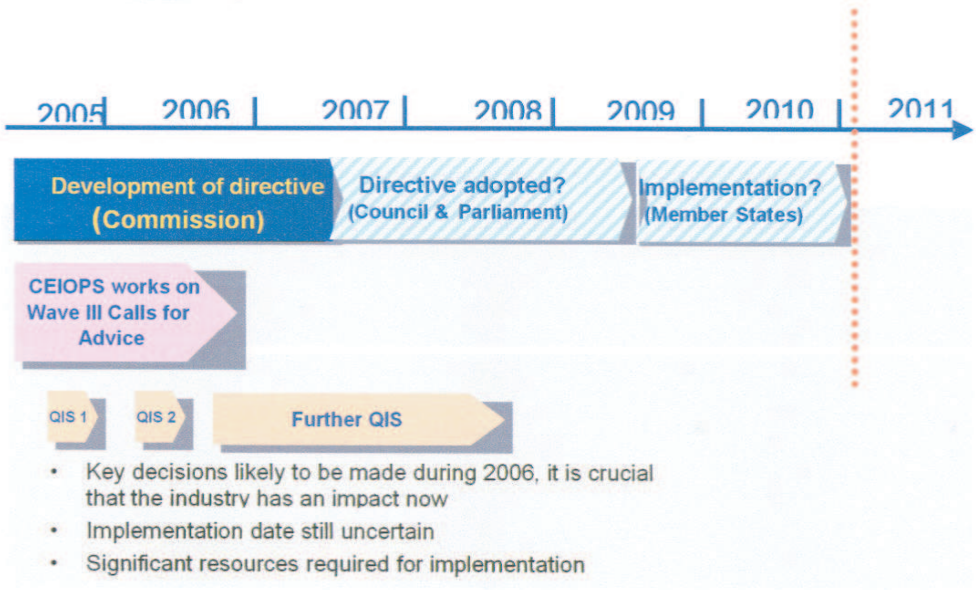


Figure 10: Solvency Time Table for the Following Years

5. WHEN DO WE HAVE TO UNDERTAKE THIS VAST PROJECT?

Figure 10 gives an indication of the timetable for the next few years:

- After examining the input of the two planned quantitative impact studies, the European Commission will prepare a directive during 2006, taking into account the advice submitted by CEIOPS.
- This preliminary text will be examined by all stakeholders and presented to the European Council and the Parliament for official adoption in 2008 (2009?).
- Member states will then have one or two years to implement the new directive.

European insurers appreciate the way in which European authorities are undertaking this vast project and calling for advice at several points during the procedure. The definition of new solvency regulations, necessary due to the inadequacy of current Solvency I regulations and the evaluation of the real risks taken by an insurer, will serve as the first positive illustration of the European Commission's new approach whereby impact studies are carried out before policy is defined.

CONTRIBUTED TALKS

ACTUARIAL STATISTICS AND MIXED MODELS: APPLICATIONS AND OPPORTUNITIES

Katrien Antonio and Jan Beirlant

University Center for Statistics, KULeuven, W. de Croylaan 54, 3001 Leuven, Belgium
Email: katrien.antonio@econ.kuleuven.be, jan.beirlant@wis.kuleuven.be

Abstract

The purpose of this paper is twofold. On the one hand, it is a short overview of our recent work on the use of mixed model methodology in actuarial statistics, which covers topics from credibility, claims reserving and non-life ratemaking. On the other hand, opportunities and challenges for future research are sketched.

1. INTRODUCTION

We discuss how mixed models can be applied in the analysis of insurance data and the decision making process following it. Starting point for the use of mixed models in actuarial statistics are traditional credibility models and their connection with linear mixed models. The credibility ratemaking problem concerns the prediction of future claims of a risk class, given past claims of that and related risk classes. Traditional credibility formulas can be reconstructed using the explicit expressions for the maximum likelihood estimations (MLE) of the fixed effects and the best linear unbiased predictor (BLUP) for the random effects in a linear mixed model. This appealing analogy was presented in Frees et al. (1999) and is a first step towards the interpretation of traditional credibility schemes in the framework of generalized linear models, using the methodology of generalized linear mixed models.

Next to the credibility ratemaking problem, examples from loss reserving and non-life ratemaking with mixed models are discussed. Using the concept of mixed models, their connection with smoothing methods and their implementation with Bayesian statistics, we present some new and promising alternatives for the techniques that are currently in use.

Section 2 contains a brief overview of the statistical concepts that are involved. In Section 3 some concrete examples are discussed and possibilities for further research are sketched. More details regarding the material presented here, are given in Antonio et al. (2006), Antonio and Beirlant (2006a) and Antonio and Beirlant (2006b).

2. STATISTICAL DETAILS

2.1. Linear mixed models (LMMs): specification and estimation

Linear mixed models extend classical linear regression models by incorporating random effects in the structure for the mean. Assume the data set at hand consists of N subjects. Let n_i denote the number of observations for subject i and $\boldsymbol{Y}_i$ its vector of observations ($1 \leq i \leq N$). The general linear mixed model is given by

$$\boldsymbol{Y}_i = \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{\epsilon}_i. \tag{1}$$

$\boldsymbol{\beta}$ ($p \times 1$) contains the parameters for the p fixed effects in the model; these are fixed, but unknown, regression parameters, common to all subjects. $\boldsymbol{b}_i$ ($q \times 1$) is the vector with the random effects for the i^{th} subject in the data set. The use of random effects reflects the belief that there is heterogeneity among subjects for a subset of the regression coefficients in $\boldsymbol{\beta}$. $\boldsymbol{X}_i$ $(n_i \times p)$ and $\boldsymbol{Z}_i$ $(n_i \times q)$ are the design matrices for the p fixed and q random effects. $\boldsymbol{\epsilon}_i$ $(n_i \times 1)$ contains the residual components for subject i. Independence between subjects is assumed. $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ are also assumed to be independent and we follow the traditional assumption that they are normally distributed with mean vector $\mathbf{0}$ and covariance matrices, say $\boldsymbol{D}$ $(q \times q)$ and $\boldsymbol{\Sigma}_i$ $(n_i \times n_i)$, respectively. Different structures for these covariance matrices are possible; an overview of some frequently used ones can be found in Verbeke and Molenberghs (2000). It is easy to see that $\boldsymbol{Y}_i$ then has a marginal normal distribution with mean $\boldsymbol{X}_i\boldsymbol{\beta}$ and covariance matrix $\boldsymbol{V}_i = \text{Var}(\boldsymbol{Y}_i)$, given by

$$\boldsymbol{V}_i = \boldsymbol{Z}_i \boldsymbol{D} \boldsymbol{Z}_i^{'} + \boldsymbol{\Sigma}_i. \tag{2}$$

In this interpretation it becomes clear that the fixed effects enter only the mean $\text{E}[Y_{ij}]$, whereas the inclusion of subject-specific effects specifies the structure of the covariance between observations on the same unit.

Denote the unknown parameters in the covariance matrix $\boldsymbol{V}_i$ with $\boldsymbol{\alpha}$. Conditional on $\boldsymbol{\alpha}$, a closed form expression for the maximum likelihood estimator of $\boldsymbol{\beta}$ exists, namely

$$\hat{\boldsymbol{\beta}} = (\sum_{i=1}^{N} \boldsymbol{X}_i^{'} \boldsymbol{V}_i^{-1} \boldsymbol{X}_i)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i^{'} \boldsymbol{V}_i^{-1} \boldsymbol{Y}_i. \tag{3}$$

To predict the random effects, the mean of the posterior distribution of the random effects given the data, $\boldsymbol{b}_i|\boldsymbol{Y}_i$, is used. Conditional on $\boldsymbol{\alpha}$, we have

$$\hat{\boldsymbol{b}}_i = \boldsymbol{D}\boldsymbol{Z}_i^{'}\boldsymbol{V}_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}), \tag{4}$$

which can be proven to be the Best Linear Unbiased Predictor (BLUP) of $\boldsymbol{b}_i$ (where 'best' is in the sense of minimal mean squared error). For estimation of $\boldsymbol{\alpha}$ maximum likelihood (ML) or restricted maximum likelihood (REML) is used. The expression maximized by the ML (L_1), respectively REML (L_2), estimates is given by

$$L_1(\boldsymbol{\alpha}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N) = c_1 - \frac{1}{2}\sum_{i=1}^{N} \log|\boldsymbol{V}_i| - \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{r}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i \tag{5}$$

$$L_2(\boldsymbol{\alpha}; \boldsymbol{y}_1, \ldots, \boldsymbol{y}_N) = c_2 - \frac{1}{2}\sum_{i=1}^{N} \log|\boldsymbol{V}_i| - \frac{1}{2}\sum_{i=1}^{N} \log|\boldsymbol{X}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{X}_i| - \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{r}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{r}_i, \tag{6}$$

where $\boldsymbol{r}_i = \boldsymbol{y}_i - \boldsymbol{X}_i \left(\sum_{i=1}^N \boldsymbol{X}_i^{'}\boldsymbol{V}_i\boldsymbol{X}_i\right)^{-1}\left(\sum_{i=1}^N \boldsymbol{X}_i^{'}\boldsymbol{V}_i^{-1}\boldsymbol{y}_i\right)$ and c_1, c_2 are appropriate constants. Equations (5) and (6) are maximized using iterative numerical techniques such as Fisher scoring or Newton-Raphson. In (3) and (4) the unknown $\boldsymbol{\alpha}$ is then replaced with $\hat{\boldsymbol{\alpha}}_{ML}$ or $\hat{\boldsymbol{\alpha}}_{REML}$, leading to the empirical BLUE for $\boldsymbol{\beta}$ and the empirical BLUP for $\boldsymbol{b}_i$. For inference regarding the fixed and random effects and the variance components, appropriate likelihood ratio and Wald tests are explained in Verbeke and Molenberghs (2000).

2.2. Generalized linear mixed models (GLMMs): specification and estimation

GLMMs extend generalized linear models (GLMs) by allowing for random, or subject-specific, effects in the linear predictor. These models are useful when the interest of the analyst lies in the individual response profiles rather than the marginal mean $\text{E}[Y_{ij}]$. The inclusion of random effects in the linear predictor reflects the idea that there is natural heterogeneity across subjects in (some of) their regression coefficients. Diggle et al. (2002) and Molenberghs and Verbeke (2005) are useful references for full details on GLMMs.

Say we have a data set at hand consisting of N subjects. For each subject i ($1 \leq i \leq N$), n_i observations are available. Given the vector $\boldsymbol{b}_i$ with the random effects for subject (or cluster) i, the repeated measurements $Y_{i1}, \ldots, Y_{in_i}$ are assumed to be independent with a density from the exponential family

$$f(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) = \exp\left(\frac{y_{ij}\theta_{ij} - \psi(\theta_{ij})}{\phi} + c(y_{ij}, \phi)\right), \; j = 1, \ldots, n_i. \tag{7}$$

Similar to a GLM, the following (conditional) relations hold

$$\mu_{ij} = \text{E}[Y_{ij}|\boldsymbol{b}_i] = \psi^{'}(\theta_{ij}) \;\text{ and }\; \text{Var}[Y_{ij}|\boldsymbol{b}_i] = \phi\psi^{''}(\theta_{ij}) = \phi V(\mu_{ij}) \tag{8}$$

where $g(\mu_{ij}) = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{'}\boldsymbol{b}_i$. As before, $g(\cdot)$ is called the link and $V(\cdot)$ the variance function. $\boldsymbol{\beta}$ ($p \times 1$) denotes the fixed effects parameter vector and $\boldsymbol{b}_i$ ($q \times 1$) the random effects vector. $\boldsymbol{x}_{ij}$ ($p \times 1$) and $\boldsymbol{z}_{ij}$ ($q \times 1$) contain subject i's covariate information for the fixed and random effects, respectively. The specification of the GLMM is completed by assuming that the random effects, $\boldsymbol{b}_i$ ($i = 1, \ldots, N$), are mutually independent and identically distributed with density function $f(\boldsymbol{b}_i|\boldsymbol{\alpha})$. Hereby $\boldsymbol{\alpha}$ denotes (again) the unknown parameters in the density. Traditionally, one works under the assumption of (multivariate) normally distributed random effects with zero mean and covariance matrix determined by $\boldsymbol{\alpha}$. Correlation between observations on the same subject arises because they share the same random effects $\boldsymbol{b}_i$.

The likelihood function for the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and ϕ then becomes (with $\boldsymbol{y} = (\boldsymbol{y}_1^{'}, \ldots, \boldsymbol{y}_N^{'})^{'}$)

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi; \boldsymbol{y}) &= \prod_{i=1}^N f(\boldsymbol{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}, \phi) \\ &= \prod_{i=1}^N \int \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) f(\boldsymbol{b}_i|\boldsymbol{\alpha}) d\boldsymbol{b}_i, \end{aligned} \tag{9}$$

where the integral is with respect to the q dimensional vector $\boldsymbol{b}_i$. When both the data and the random effects are normally distributed (as in the linear mixed model), the integral can be worked out analytically and closed-form expressions exist for the maximum likelihood estimator of $\boldsymbol{\beta}$ and the BLUP for $\boldsymbol{b}_i$ (see (3) and (4)). For general GLMMs, however, approximations to the likelihood or numerical integration techniques are required to maximize equation (9) with respect to the unknown parameters. Restricted pseudo-likelihood ((RE)PL) (Wolfinger and O'Connell (1993)) and (adaptive) Gauss-Hermite quadrature (Liu and Pierce (1994)) are two widely used techniques to perform the maximum likelihood estimation. Both techniques are available in the commercial software package SAS and their use will be illustrated in Section 3. The pseudo-likelihood technique corresponds with the penalized quasi-likelihood (PQL) method of Breslow and Clayton (1993). Since maximum likelihood techniques are hindered by the integration over the q-dimensional vector of random effects, a Bayesian implementation of GLMMs is considered as well. Hereby random numbers are drawn from the relevant posterior and predictive distributions using Markov Chain Monte Carlo (MCMC) techniques. WINBUGS allows easy implementation of these models. Illustrative code for both SAS and WINBUGS is available on the web [1].

2.3. Smoothing with mixed models

To provide some background for smoothing with mixed model methodology, let us start from the simple example of scatterplot smoothing. Data (x_i, y_i) $(i = 1, \ldots, n)$ are given and the model $Y_i = f(x_i) + \epsilon_i$ $(i = 1, \ldots, n)$ is fitted. To estimate the unknown function $f(\cdot)$, a linear combination of some basis functions is used. Possible basis functions are *truncated power basis functions*, *B-splines* or *radial basis functions*, among others. For truncated power basis functions of degree p with K knots $\kappa_1, \ldots, \kappa_K$ [2], define the design matrix $\boldsymbol{B}$ as

$$\boldsymbol{B} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^p & (x_1 - \kappa_1)_+^p & \ldots & (x_1 - \kappa_K)_+^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^p & (x_n - \kappa_1)_+^p & \ldots & (x_n - \kappa_K)_+^p \end{bmatrix}. \tag{10}$$

The unknown function $f(\cdot)$ is then estimated as $\widehat{f}(x) = \boldsymbol{B}(x)\widehat{\boldsymbol{\beta}}$ where $\boldsymbol{B}(x)$ is a row vector, similar to a row from $\boldsymbol{B}$, and $\widehat{\boldsymbol{\beta}}$ is the solution of the least-squares problem $\min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - \boldsymbol{B}(x_i)\boldsymbol{\beta})^2$, subject to the constraint $\sum_{k=1}^K \beta_{pk}^2 < C$ to obtain a smooth fit. Hereby, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p, \beta_{p1}, \ldots, \beta_{pK})'$ and thus the penalized coefficients correspond with the truncated power functions. Using a Lagrange multiplier argument, this optimization problem is rewritten as

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^n (y_i - \boldsymbol{B}(x_i)\boldsymbol{\beta})^2 + \alpha \boldsymbol{\beta}' \boldsymbol{P} \boldsymbol{\beta}, \tag{11}$$

where α is the so-called smoothing parameter and $\boldsymbol{P}$ a penalty matrix given by

$$\boldsymbol{P} = \begin{bmatrix} 0_{p+1\times p+1} & 0_{p+1\times K} \\ 0_{K\times p+1} & \mathbf{I}_{K\times K} \end{bmatrix}. \tag{12}$$

[1] see http://www.econ.kuleuven.be/katrien.antonio

[2] The truncated line $(x - \kappa_k)_+$ is zero, when $x < \kappa_k$ and equals $x - \kappa_k$ elsewhere. $(x - \kappa_k)_+^p$ has to be interpreted as $\{(x - \kappa_k)_+\}^p$. The basis functions $\{1, x, x^2, \ldots, x^p, (x - \kappa_1)_+^p, \ldots, (x - \kappa_K)_+^p\}$ span the vector space of piecewise functions of degree p with knots at $\kappa_1, \ldots, \kappa_K$.

Ruppert et al. (2003) (among others) rewrite the argument of the optimization problem in (11), after dividing by σ_ϵ^2, as

$$\frac{1}{\sigma_\epsilon^2}\|\boldsymbol{y}-\boldsymbol{X\beta}-\boldsymbol{Zu}\|^2+\frac{1}{\sigma_u^2}\|\boldsymbol{u}\|^2, \tag{13}$$

where $\sigma_u^2=\sigma_\epsilon^2/\alpha$, $\boldsymbol{y}=(y_1,\ldots,y_n)^{'}$, $\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_p)^{'}$ (i.e. the regression parameters for the basis functions $1,\ x,\ x^2,\ldots,\ x^p$), $\boldsymbol{u}=(\beta_{p1},\ldots,\beta_{pK})^{'}$,

$$\boldsymbol{X}=\begin{bmatrix}1 & x_1 & x_1^2 & \ldots & x_1^p\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_n & x_n^2 & \ldots & x_n^p\end{bmatrix} \text{ and } \boldsymbol{Z}=\begin{bmatrix}(x_1-\kappa_1)_+^p & \ldots & (x_1-\kappa_K)_+^p\\ \vdots & \vdots & \vdots\\ (x_n-\kappa_1)_+^p & \ldots & (x_n-\kappa_K)_+^p\end{bmatrix}. \tag{14}$$

By considering $\boldsymbol{u}$ as random effects with $\boldsymbol{u}\sim N(0,\sigma_u^2\boldsymbol{I}_{K\times K})$, (13) reduces to minus two times the log-likelihood of $(\boldsymbol{Y},\boldsymbol{u})$ in the linear mixed model $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{Zu}+\boldsymbol{\epsilon}$, under the assumptions $\boldsymbol{Y}|\boldsymbol{u}\sim N(\boldsymbol{X\beta}+\boldsymbol{Zu},\sigma_\epsilon^2\boldsymbol{I})$, $\boldsymbol{u}\sim N(\boldsymbol{0},\sigma_u^2\boldsymbol{I})$ and $\boldsymbol{\epsilon}\sim N(\boldsymbol{0},\sigma_\epsilon^2\boldsymbol{I})$.

A similar reasoning leads to the penalized splines formulation of a GAM, where $Y_1,\ldots,Y_n$ are independent random variables with a density $f(\cdot)$ from the exponential family and an additive predictor $\eta_i=\sum_{h=1}^{l}f_h(x_{ih})$ $(i=1,\ldots,n)$. Construct the design matrix $\boldsymbol{X}$ as

$$\boldsymbol{X}=\left[\begin{array}{ccccc|c|cccc}1 & x_{11} & x_{11}^2 & \ldots & x_{11}^p & \ldots & x_{1l} & x_{1l}^2 & \ldots & x_{1l}^p\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & x_{n1} & x_{n1}^2 & \ldots & x_{n1}^p & \ldots & x_{nl} & x_{nl}^2 & \vdots & x_{nl}^p\end{array}\right]. \tag{15}$$

In the above specification the l blocks specify the unpenalized basis functions for estimation of the unknown functions $f_1(\cdot),\ldots,f_l(\cdot)$. As in the scatterplot smoothing example, a smooth fit results by putting constraints on the coefficients of the truncated basis functions. This is done by treating them as random effects in a mixed model formulation. Define

$$\mathbf{Z}^{pen}=\left[\begin{array}{ccc|c|ccc}(x_{11}-\kappa_1^1)_+^p & \ldots & (x_{11}-\kappa_{K_1}^1)_+^p & \ldots & (x_{1l}-\kappa_1^l)_+^p & \ldots & (x_{1l}-\kappa_{K_l}^l)_+^p\\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\ (x_{n1}-\kappa_1^1)_+^p & \ldots & (x_{n1}-\kappa_{K_1}^1)_+^p & \ldots & (x_{nl}-\kappa_1^l)_+^p & \ldots & (x_{nl}-\kappa_{K_l}^l)_+^p\end{array}\right], \tag{16}$$

where K_i denotes the number of knots to estimate $f_i(\cdot)$ $(i=1,\ldots,l)$. In case of a GAM, the log-likelihood is considered as a function of the additive predictor $\boldsymbol{\eta}$ and, using penalized regression splines, $\widehat{\boldsymbol{\eta}}=\boldsymbol{X}\widehat{\boldsymbol{\beta}}+\boldsymbol{Z}\widehat{\boldsymbol{u}}$, where $\widehat{\boldsymbol{\beta}}$ is obtained from the following penalized log-likelihood

$$\max_{\boldsymbol{\beta}}\{\boldsymbol{y}^{'}(\boldsymbol{X\beta}+\boldsymbol{Zu})-\boldsymbol{1}^{'}\psi(\boldsymbol{X\beta}+\boldsymbol{Zu})\}-\frac{1}{2}\sum_{j=1}^{l}\alpha_j\boldsymbol{u}_j^{'}\boldsymbol{u}_j, \tag{17}$$

and $\widehat{\boldsymbol{u}}$ from $E[\boldsymbol{u}|\boldsymbol{y}]$ where – for ease of notation – a canonical link is assumed. $\boldsymbol{\beta}$ is the column vector with the parameters for the unpenalized basis functions in (15) (one parameter per column of $\boldsymbol{X}$). $\boldsymbol{u}_j=(u_{j1},\ldots,u_{jK_j})^{'}$ $(j=1,\ldots,l)$, α_j $(j=1,\ldots,l)$ is the smoothing parameter for function $f_j(\cdot)$ and say $\boldsymbol{u}=(\boldsymbol{u}_1^{'},\ldots,\boldsymbol{u}_l^{'})^{'}$. The optimization problem in (17) is equivalent to the optimization problem in a generalized linear mixed model (see Breslow and Clayton (1993)) with the GLMM specified as

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{u}) &= \exp\left(\boldsymbol{y}^{'}(\boldsymbol{X\beta}+\boldsymbol{Zu}) - \mathbf{1}^{'}\psi(\boldsymbol{X\beta}+\boldsymbol{Zu}) + \mathbf{1}^{'}c(\boldsymbol{y})\right), \\ \boldsymbol{u} &\sim N(\mathbf{0}, \boldsymbol{\Lambda}), \\ \text{and } \boldsymbol{\Lambda} &= \begin{bmatrix} \sigma_1^2\mathbf{I}_{K_1\times K_1} & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_l^2\mathbf{I}_{K_l\times K_l} \end{bmatrix}, \end{aligned} \tag{18}$$

where $\sigma_j^2 = 1/\alpha_j$ $(j = 1, \ldots, l)$ and – again – a canonical link is used in (18) for ease of notation. Both (17) and (18) are easily generalized to the case of a non-canonical link.

In line with the previous specifications, a GAMM for longitudinal data can be rewritten as a GLMM as well. Let Y_{ij} denote the j^{th} observation for subject i, where $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$. Conditional on the random effects $\boldsymbol{b}_i$ $(q \times 1)$ for subject i (and $\boldsymbol{b}_i \sim N(\mathbf{0}, \boldsymbol{D})$), $Y_{i1}, \ldots, Y_{in_i}$ are independent with a density from the exponential family and a predictor $\eta_{ij} = \sum_{h=1}^{l} f_h(x_{ijh}) + \boldsymbol{z}_{ij}^{'}\boldsymbol{b}_i$. Specify the design matrices $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ for subject i $(i = 1, \ldots, N)$ as

$$\boldsymbol{X}_i = \left[\begin{array}{ccccc|c|cccc} 1 & x_{i11} & x_{i11}^2 & \ldots & x_{i11}^p & \ldots & x_{i1l} & x_{i1l}^2 & \ldots & x_{i1l}^p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{in_i1} & x_{in_i1}^2 & \ldots & x_{in_i1}^p & \ldots & x_{in_il} & x_{in_il}^2 & \vdots & x_{in_il}^p \end{array}\right], \tag{19}$$

and

$$\mathbf{Z}_i^{pen} = \left[\begin{array}{ccc|c|ccc} (x_{i11}-\kappa_1^1)_+^p & \ldots & (x_{i11}-\kappa_{K_1}^1)_+^p & \ldots & (x_{i1l}-\kappa_1^l)_+^p & \ldots & (x_{i1l}-\kappa_{K_l}^l)_+^p \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (x_{in_i1}-\kappa_1^1)_+^p & \ldots & (x_{in_i1}-\kappa_{K_1}^1)_+^p & \ldots & (x_{in_il}-\kappa_1^l)_+^p & \ldots & (x_{in_il}-\kappa_{K_l}^l)_+^p \end{array}\right]. \tag{20}$$

Together with the ‘classical’ design matrix for the random effects for $\boldsymbol{b}_i$ $(i = 1, \ldots, N)$,

$$\mathbf{Z}_i^{ran} = \begin{bmatrix} z_{i11} & \ldots & z_{i1q} \\ \vdots & \ddots & \vdots \\ z_{in_i1} & \ldots & z_{in_iq} \end{bmatrix} \text{ and } \mathbf{Z}_i = [\mathbf{Z}_i^{pen}|\mathbf{Z}_i^{ran}], \tag{21}$$

the contribution of subject i to the GLMM specification of the GAMM is given by

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{r}_i) &= \exp\left(\boldsymbol{y}_i^{'}(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{r}_i) - \mathbf{1}^{'}\psi(\boldsymbol{X}_i\boldsymbol{\beta}+\boldsymbol{Z}_i\boldsymbol{r}_i) + \mathbf{1}^{'}c(\boldsymbol{y}_i)\right), \\ \boldsymbol{r}_i &= (\boldsymbol{u}^{'}, \boldsymbol{b}_i^{'})^{'} \sim N(\mathbf{0}, \boldsymbol{\Lambda}_i), \\ \text{and } \boldsymbol{\Lambda}_i &= \begin{bmatrix} \sigma_1^2\mathbf{I}_{K_1\times K_1} & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma_l^2\mathbf{I}_{K_l\times K_l} & 0 \\ 0 & 0 & \ldots & 0 & \boldsymbol{D} \end{bmatrix}. \end{aligned} \tag{22}$$

The assumption of independence among subjects completes the specification of the GLMM representation of the GAMM.

3. APPLICATIONS AND OPPORTUNITIES

3.1. Credibility

Using linear mixed models Frees et al. (1999) already gave a longitudinal data analysis interpretation of the well-known credibility models of Bühlmann (1967), Bühlmann (1969), Bühlmann and Straub (1970), Hachemeister (1975) and Jewell (1975). They explained how to specify the fixed and random effects for every subject or risk class i ($i = 1, \ldots, N$) and used $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}_i$ (as in (3) and (4)) to derive the Best Linear Unbiased Predictor for the conditional mean of a future observation ($\text{E}[Y_{i,n_i+1}|\boldsymbol{b}_i]$). For the above mentioned credibility models, this BLUP corresponds with the classical credibility formulas.

However, the normal-normal model (normality for both responses and random effects) will not always be plausible for the data at hand (which can be, for instance, counts, binary or skewed data). Therefore it is useful to revisit the credibility models in the context of GLMs and to consider their specification as a GLMM. In this way, estimators and predictors will be used that take the distributional features of the data into account.

Interpreting traditional credibility models in the context of GLMMs implies that the additive regression structure in terms of fixed and subject-specific (or risk class specific) effects is specified on the scale of the linear predictor, namely

$$g(\mu_{ij}) = \eta_{ij} = \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta} + \boldsymbol{z}_{ij}^{'}\boldsymbol{b}_i. \tag{23}$$

Hereby i ($i = 1, \ldots, N$) denotes the subject, for instance a policy(holder) or risk class, and j refers to its j^{th} measurement, unless it is stated otherwise. The link function $g(\cdot)$ and variance function $V(\cdot)$ are determined by the chosen GLM. More details are given in Antonio and Beirlant (2006a).

3.2. Claims reserving

We illustrate how information on claim counts and claim amounts can be combined in a semiparametric regression model for claims reserving. Using a Bayesian implementation of the smoothers from Section 2, the data considered in de Alba (2002) are reanalyzed. A generalized additive model is constructed that combines data on claim numbers and claim intensities. We illustrate that, by using Bayesian statistics, simulation from the predictive distributions in this more complicated model is possible without many additional efforts. Full details are in Antonio and Beirlant (2006b).

Denote with Y_{ij} the aggregate payment for cell (i, j) and let N_{ij} be the corresponding number of claims. Thus, $Y_{ij} = \sum_{k=1}^{N_{ij}} Y_{ijk}$, with Y_{ijk} the payments composing the aggregate claim Y_{ij}. Following de Alba (2002), a model is considered which combines information on the number of claims registered and the total amount paid out for these claims, per arrival/development year combination. Let $Z_{ij} := Y_{ij}/N_{ij}$ be the average payment for cell (i,j) and model

$$\begin{aligned}
Z_{ij} &\sim \Gamma(\nu, \mu_{ij}^{Av}/\nu), \\
\text{where } \log(\mu_{ij}^{Av}) &= \alpha_1^{Av} * I(i=1) + \cdots + \alpha_{10}^{Av} * I(i=10) + f^{Av}(j) \\
\text{and } \frac{N_{ij}}{\phi} &\sim \text{Poisson}\left(\frac{\mu_{ij}^{Num}}{\phi}\right), \\
\text{where } \log(\mu_{ij}^{Num}) &= \alpha_1^{Num} * I(i=1) + \cdots + \alpha_{10}^{Num} * I(i=10) + f^{Num}(j). \qquad (24)
\end{aligned}$$

Furthermore, the Z_{ij}'s and N_{ij}'s are assumed to be independent.

Based on an inspection of the scatterplots and residual plots from an analysis with Proc Glimmix in SAS (not shown), 4 knots in the direction of development years, with positions $(2,3,5,7)$ (for claim counts and average payments), are used. Results for the reserves from this model are summarized in Table 1 (claim counts) and Table 2 (total payments, obtained by multiplying claim numbers and average payments).

	Mean Poisson	Mean o-Poisson	St.Dev. Bayes.	5% Bayes.	50% Bayes.	97.5% Bayes.
AY 2	2	2	4.36	0	0	17
AY 3	7	5	7.424	0	0	25
AY 4	13	9	10.372	0	8	34
AY 5	22	19	14.418	0	17	51
AY 6	41	40	21.06	8	34	85
AY 7	97	96	33.702	34	93	169
AY 8	149	147	47.275	68	144	246
AY 9	240	240	84.071	102	229	432
AY 10	332	322	215.339	42	279	855
Total	902	879	248.871	500	847	1,465

Table 1: *Predictive distribution for the number of claims: results from a Bayesian analysis with truncated line basis functions for smooth function over development years. A burn-in of 50,000 simulations was used, followed by another 450,000 simulations to which a thinning factor of 10 was applied.*

3.3. Non-life ratemaking

We consider a data set from Frees et al. (2001). These authors focused on the longitudinal character of the data and modelled the logarithmic transformation of 'PP=Loss/Payroll', using linear mixed models. Our analysis as well takes the longitudinal character of the data into account and considers inference and prediction regarding individual risk classes. Use is made, however, of a gamma GLMM; in this way no transformation of the data is required. 'Loss' is the response variable and

	Mean Bayes.	St.Dev. Bayes.	2.5% Bayes.	50% Bayes.	97.5% Bayes.
AY 2	165	500	0	0	2
AY 3	372	742	0	0	2
AY 4	606	909	0	312	3
AY 5	1,038	1,127	0	726	3,963
AY 6	1,562	1,306	111	1,239	4,908
AY 7	2,473	1,612	523	2,103	6,510
AY 8	3,802	2,328	947	3,288	9,694
AY 9	5,503	3,522	1,344	4,673	14,507
AY 10	5,983	5,937	495	4,242	21,772
Total	21,503	8,990	9,513	19,753	43,903

Table 2: *Predictive distribution of the reserves (data displayed in thousands): results from a Bayesian analysis with truncated line basis functions for smooth functions over development period. A burn-in of 50,000 simulations was used, followed by another 450,000 simulations to which a thinning factor of 10 was applied.*

'Payroll' is used as an offset. The following models are considered

$$\begin{aligned} Y_{ij}|\boldsymbol{b}_i &\sim \Gamma(\nu, \mu_{ij}/\nu) \\ \text{where } \log(\mu_{ij}) &= \log(Payroll_{ij}) + \beta_0 + b_{i,0} \quad (25) \\ \text{versus } \log(\mu_{ij}) &= \log(Payroll_{ij}) + \beta_0 + \beta_1 Year_{ij} + b_{i,0} \quad (26) \\ \text{and } \log(\mu_{ij}) &= \log(Payroll_{ij}) + \beta_0 + \beta_1 Year_{ij} + b_{i,0} + b_{i,1} Year_{ij}. \quad (27) \end{aligned}$$

The gamma density function is specified as $f(y) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu y}{\mu}\right)^{\nu} \exp\left(\frac{-\nu y}{\mu}\right)\frac{1}{y}$. The specification in (27) did not lead to convergence of the SAS procedures. Structure (26) is the preferred choice for the linear predictor. Table 3 contains the results of a maximum-likelihood and Bayesian analysis, where non-informative priors were used. Fitted values against real observations are plotted in Figure 1. More details and related examples are in Antonio and Beirlant (2006a).

	PQL		*adaptive G-H*		*Bayesian*	
	Est.	SE	Est.	SE	Mean	90% Cred. Int.
β_0	-4.172	0.091	-4.148	0.091	-4.147	(-4.298, -3.996)
β_1	0.042	0.012	0.042	0.012	0.042	(0.022, 0.062)
δ_1	0.915	0.128	0.912	0.127	0.938	(0.741, 1.174)

Table 3: *Workers' compensation data (Losses): results of maximum likelihood and Bayesian analysis. REML is used in PQL.*

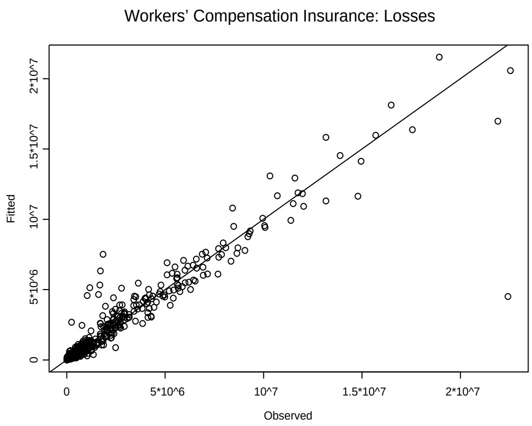


Figure 1: *Workers' compensation data (Losses): observed versus fitted values.*

3.4. Discussion

We presented some new statistical approaches for the analysis of actuarial data related to claims reserving and credibility. To illustrate further possibilities in this framework, we mention three interesting topics of our current research. Firstly, it is interesting to compare the mixed model approach with a copula construction to model the dynamics in panel data (as in Frees and Wang (2005)). Secondly, the joint modelling of longitudinal data on claim numbers and claim amounts through a mixed model, can be considered and contrasted with – again – a copula construction. Thirdly, instead of working in the framework of the exponential distribution, regression models for heavy-tailed data are of interest for actuaries. In this way, a combination of the models discussed above with heavy-tailed regression models, can be useful for actuarial applications.

References

K. Antonio and J. Beirlant. Actuarial statistics with generalized linear mixed models. *Insurance: Mathematics and Economics*, 2006a. to be published.

K. Antonio and J. Beirlant. Semiparametric regression models for claims reserving and credibility: the mixed model approach. *Working Paper (online at www.econ.kuleuven.be/katrien.antonio)*, 2006b.

K. Antonio, J. Beirlant, T. Hoedemakers, and R. Verlaak. Lognormal mixed models for reported claims reserves. *North American Actuarial Journal*, 10:30–48, 2006.

N.E. Breslow and D.G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Society*, 88(421):9–25, 1993.

H. Bühlmann. Experience rating and credibility I. *ASTIN Bulletin*, 4:199–207, 1967.

H. Bühlmann. Experience rating and credibility II. *ASTIN Bulletin*, 5:157–165, 1969.

H. Bühlmann and E. Straub. Glaubwürdigkeit für Schadensätze. *Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker*, pages 111–133, 1970.

E. de Alba. Bayesian estimation of outstanding claims reserves. *North American Actuarial Journal*, 6(4):1–20, 2002.

P. Diggle, P. Heagerty, K.Y. Liang, and S.L. Zeger. *Analysis of longitudinal data.* Oxford University Press, Oxford, 2002.

E.W. Frees and P. Wang. Credibility using copulas. *North American Actuarial Journal*, 9(2): 31–48, 2005.

E.W. Frees, V.R. Young, and Y. Luo. A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics*, 24:229–247, 1999.

E.W. Frees, V.R. Young, and Y. Luo. Case studies using panel data models. *North American Actuarial Journal*, 5(4):24–42, 2001.

C.A. Hachemeister. Credibility for regression models with application to trend. In *Proceedings of the Berkeley Actuarial Research Conference on Credibility*, pages 129–163, New York, 1975. Adamic Press.

W.S. Jewell. The use of collateral data in credibility theory. *Giornale dell' Instituto Italiano degli Actuary*, 38:1–16, 1975.

Q. Liu and D.A. Pierce. A note on Gauss-Hermite quadrature. *Biometrika*, 81(3):624–629, 1994.

G. Molenberghs and G. Verbeke. *Models for discrete longitudinal data.* Springer series in statistics. Springer, New York, 2005.

D. Ruppert, M.P. Wand, and R.J. Carroll. *Semiparametric regression.* Cambridge University Press, Cambridge, 2003.

G. Verbeke and G. Molenberghs. *Linear mixed models for longitudinal data.* Springer, New York, 2000.

R. Wolfinger and M. O'Connell. Generalized linear mixed models: a pseudo-likelihood approach. *Journal of Statistical Computation and Simulation*, 48:233–243, 1993.

POLICY ITERATION METHOD FOR AMERICAN OPTIONS [1]

Christian Bender†, Anastasia Kolodko†,§ and John Schoenmakers†

†*Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany*
§*Institute of Computational Mathematics and Mathematical Geophysics, Russian Acad. Science, Lavrentieva str. 6, 630090 Novosibirsk, Russia*
Email: bender[kolodko][schoenma]@wias-berlin.de

Abstract

This paper is an overview of recent results by Kolodko and Schoenmakers (2006), Bender and Schoenmakers (2006) on the evaluation of options with early exercise opportunities via policy improvement. Stability is discussed and simulation results based on plain Monte Carlo estimators for conditional expectations are presented.

1. INTRODUCTION

The evaluation of American style derivatives on a high dimensional underlying is an important and challenging problem. Typically these derivatives cannot be priced by the classical PDE methods, as the computational cost rapidly increases with the dimension of the underlying. This problem is known as the 'curse of dimensionality'. Only in recent years several approaches have been proposed to overcome this problem. These methods basically rely on Monte Carlo simulation and can be roughly divided into three groups. The first group directly employs a recursive scheme for solving the stopping problem, known as backward dynamic programming. Different techniques are applied to approximate the nested conditional expectations. The stochastic mesh method by Broadie et al. (2000) and the least square regression method of Longstaff and Schwartz (2001) are among the most popular approaches in this group. An alternative to backward dynamic programming is to approximate the exercise boundary by simulation, see e.g. Andersen (1999), Ibáñez and Zapatero (2004). The third group relies on a dual approach developed in Rogers (2002), Haugh and Kogan (2004), and in a multiplicative setting by Jamshidian (1997). For a numerical treatment of this approach, see Kolodko and Schoenmakers (2004). By duality, tight price upper bounds may be constructed from given approximative processes.

In this paper we survey a new policy iteration for discretized American options which was recently introduced in Kolodko and Schoenmakers (2006) and Bender and Schoenmakers (2006).

[1] Supported by the DFG Research Center MATHEON 'Mathematics for Key Technologies' in Berlin.

The method is mending one of main drawbacks of backward dynamic programming: Suppose exercise can take place at one out of k time instances. Then, in order to obtain the value of the optimal stopping problem via backward dynamic programming, one has to calculate nested conditional expectations of order k. No approximation of the time 0 value is available prior to the evaluation of the kth nested conditional expectations. This prevents the use of plain Monte Carlo simulations for approximating the conditional expectations and requires more complicated approximation procedures for these quantities. For instance, to employ the procedure of Longstaff and Schwartz (2001), one has to choose the number of basis functions and the basis functions themselves, i.e. the approximation procedure must be differently tailored to different derivatives. Contrary, our policy iteration yields approximations of the time 0 value of the value function for every iteration step, which monotonically increase to the Snell envelope. This allows to calculate some approximations of the Snell envelope by plain Monte Carlo simulations. The algorithm converges in the same number of steps as backward dynamic programming does. So theoretically, the algorithm is as good as backward dynamic programming.

After recalling the optimal stopping problem in section 2, we introduce our policy iteration in section 3.1. Note, the policy iteration is different from Howard (1960) policy iteration for backward dynamic programming and can be shown to yield better approximations. Stability of the policy improvement is discussed in section 3.2. It turns out, that the shortfall of the perturbed policy improvement under the theoretical policy improvement converges to zero. Surprisingly, the distance need not convergence, so that the perturbed improvement can even perform better than the theoretical. Section 4 is devoted to simulations. We evaluate the price of basket-put and maximum-call on five assets, which has become a benchmark problem in recent years. The examples show that tight approximations of the option prices can be achieved with a plain Monte Carlo simulation.

2. OPTIMAL STOPPING IN DISCRETE TIME

It is well known that by the no arbitrage principle the pricing of American options is equivalent to the optimal stopping problem of the discounted derivative under a pricing measure. We now recall some facts about the optimal stopping problem in discrete time.

Suppose $(Z(i)$: $i = 0, 1, \ldots, k)$ is a nonnegative stochastic process in discrete time on a probability space $(\Omega, \mathcal{F}, P)$ adapted to some filtration $(\mathcal{F}_i : 0 \leq i \leq k)$ which satisfies

$$\sum_{i=1}^{k} E|Z(i)| < \infty.$$

We may think of the process Z as a cashflow, which an investor may exercise once. The investors' problem is to maximize his expected gain by choosing the optimal time for exercising. This problem is known as optimal stopping in discrete time.

To formalize the stopping problem we define $\mathcal{S}_i$ as the set of $\mathcal{F}_i$ stopping times taking values in $\{i, \ldots, k\}$. The stopping problem can now be stated as follows:

Find stopping times $\tau^*(i) \in \mathcal{S}_i$ such that for $0 \leq i \leq k$

$$E^{\mathcal{F}_i}\left[Z(\tau^*(i))\right] = \text{esssup}_{\tau \in \mathcal{S}_i} E^{\mathcal{F}_i}\left[Z(\tau)\right]. \tag{1}$$

The process on the right hand side is called the *Snell envelope* of Z and we denote it by $Y^*(i)$.

We collect some facts, which can be found in Neveu (1975) for example.

1. The Snell envelope Y^* of Z is the smallest supermartingale that dominates Z. It can be constructed recursively by backward dynamic programming:

$$\begin{aligned} Y^*(k) &= Z(k) \\ Y^*(i) &= \max\{Z(i),\ E^{\mathcal{F}_i}[Y^*(i+1)]\}. \end{aligned}$$

2. A family of optimal stopping times is given by

$$\widetilde{\tau}^*(i) = \inf\{i \leq j \leq k :\ Z(j) \geq Y^*(j)\}.$$

If several optimal stopping families exist, then the above family is the family of first optimal stopping times. In that case

$$\widehat{\tau}^*(i) = \inf\{i \leq j \leq k :\ Z(j) > Y^*(j)\}$$

is the family of last optimal stopping times.

3. THE POLICY ITERATION

3.1. Definition of the improvement procedure

Suppose the buyer of the option chooses ad hoc a family of stopping times $(\tau(i) :\ 0 \leq i \leq k)$ taking values in the set $\{0, \ldots, k\}$. We interpret $\tau(i)$ as the time, at which the buyer will exercise his option, provided he has not exercised prior to time i. This interpretation requires the following consistency condition:

Definition 3.1 *A family of integer-valued stopping times* $(\tau(i) :\ 0 \leq i \leq k)$ *is said to be* consistent, *if*

$$\begin{gathered} i \leq \tau(i) \leq k,\ \tau(k) \equiv k, \\ \tau(i) > i \Rightarrow \tau(i) = \tau(i+1), \qquad 0 \leq i < k. \end{gathered}$$

Indeed, suppose $\tau(i) > i$, i.e. according to our interpretation the investor has not exercised the first right prior to time $i+1$. Then he has not exercised the first right prior to time i, either. This means he will exercise the first right at times $\tau(i)$ and $\tau(i+1)$, which requires $\tau(i) = \tau(i+1)$. A typical example of a consistent stopping family can be obtained by comparison with the still-alive European options

$$\tau(i) := \inf\left\{j :\ i \leq j \leq k,\ Z(j) \geq \max_{j+1 \leq p \leq k} E^{\mathcal{F}_j}\left[Z(p)\right]\right\}. \tag{2}$$

Given some consistent stopping family τ we define a new stopping family by

$$\widetilde{\tau}(i) := \inf\left\{j :\ i \leq j \leq k,\ Z(j) \geq \max_{j+1 \leq p \leq k} E^{\mathcal{F}_j}\left[Z(\tau(p))\right]\right\}. \tag{3}$$

Note, the stopping family $\widetilde{\tau}$ is consistent. In particular $\widetilde{\tau}(k) = k$, since $\max \emptyset = -\infty$. We call $\widetilde{\tau}$ a *one-step improvement* of τ for the following reason: denote by $Y(i;\tau)$ the value process corresponding to the stopping family τ, namely

$$Y(i;\tau) = E^{\mathcal{F}_i}\left[Z(\tau(i))\right].$$

Then the one-step improvement yields a higher value than the given family,

$$Y(i;\widetilde{\tau}) \geq Y(i;\tau).$$

This will be proved in theorem 3.2 below. We note that, for example, the stopping family based on the maximum of still alive Europeans in (2) is the one-step improvement of the trivial stopping family $\tau(i) = i$.

It is natural to iterate this policy improvement: suppose τ_0 is some consistent stopping family. Define, recursively,

$$\begin{aligned} \tau_m &= \widetilde{\tau}_{m-1} \\ Y_m(i) &= Y(i;\tau_m). \end{aligned}$$

It can be shown that $Y_m(i)$ coincides with the time i value of the Snell envelope when $m \geq k - i$. This means the policy improvement algorithm is theoretically as good as backward dynamic programming, but admits to calculate increasing approximations of the Snell envelope at every iteration step.

Remark 3.1 *Given a consistent stopping family τ, $Y(0;\tau)$ is always a lower bound of $Y^*(0)$. From this lower bound an upper bound can be constructed by a duality method developed by Rogers (2002) and Haugh and Kogan (2004). Define,*

$$Y_{up}(\tau) = E\left[\max_{0\leq j\leq k}(Z(j) - M(j))\right], \tag{4}$$

where $M(0) = 0$ and, for $1 \leq i \leq k$,

$$M(i) = \sum_{p=1}^{i}\left(Y(p;\tau) - E^{\mathcal{F}_{p-1}}\left[Y(p;\tau)\right]\right).$$

Remark 3.2 *When τ^* is some optimal stopping family, the supermartingale property of the Snell envelope yields,*

$$\max_{i+1\leq p\leq k} E^{\mathcal{F}_i}\left[Z(\tau^*(p))\right] = E^{\mathcal{F}_i}\left[Y^*(i+1)\right].$$

Thus, the one-step improvement of τ^* is the family of first optimal stopping times. This shows, the latter family is the only fixed point of the one-step improvement.

3.2. Stability

In practice, we cannot expect to know analytical expressions of the conditional expectations on the right hand side of the exercise criterion in (3), but can only calculate approximations. Therefore, a stability result is called for.

Given a consistent stopping family τ and a sequence of $\mathcal{F}_i$-adapted processes $\epsilon^{(N)}(i)$ define

$$\widetilde{\tau}^{(N)}(i) := \inf \left\{ j :\ i \leq j \leq k,\ Z(j) \geq \max_{j+1 \leq p \leq k} E^{\mathcal{F}_j}\left[Z(\tau(p))\right] + \epsilon^{(N)}(j) \right\}.$$

The sequence $\epsilon^{(N)}$ accounts for the errors when approximating the conditional expectation. We suppose throughout this section that

$$\lim_{N \to \infty} \epsilon^{(N)}(i) = 0, \quad P\text{-}a.s.$$

We will first show by some simple examples that we must neither expect

$$\widetilde{\tau}^{(N)}(i) \to \widetilde{\tau}(i) \quad \text{in probability}$$

nor

$$Y(0; \widetilde{\tau}^{(N)}) \to Y(0; \widetilde{\tau}).$$

Example (i) Suppose ξ_N is a sequence of independent binary trials with $P(\xi_N = 1) = P(\xi_N = 0) = 1/2$. We define the process $(Z(i) :\ i = 0, 1)$ by $Z(0) = Z(1) \equiv 1$. The σ-field $\mathcal{F}_0 = \mathcal{F}_1$ is the one generated by the sequence of trials. Moreover, the sequence of perturbations is defined by $\epsilon^{(N)}(0) = \xi_N / N$ and $\epsilon^{(N)}(1) = 0$. Then, starting with any consistent stopping family τ, we get

$$\widetilde{\tau}^{(N)}(0) = \xi_N.$$

In particular, no subsequence of $\widetilde{\tau}^{(N)}(0)$ converges in probability.

(ii) Let $\Omega = \{\omega_0, \omega_1\}$, $\mathcal{F}$ the powerset of Ω and $P(\{\omega_1\}) = 1/4 = 1 - P(\{\omega_0\})$. We define the process $(Z(i) :\ i = 0, 1, 2)$ by $Z(0) = Z(2) = 2$ and $Z(1, \omega_0) = 1$, $Z(1, \omega_1) = 3$. $\mathcal{F}_i$ is the filtration generated by Z. We start with the stopping family $\tau(i) = i$. As $E[Z(1)] = 3/2$, we have

$$Z(0) = 2 \geq \max\{3/2,\ 2\} = \max\{E[Z(1)],\ E[Z(2)]\} = \widehat{Y}(0, \tau).$$

Therefore,

$$\widetilde{\tau}(0) = 0$$

and

$$Y(0; \widetilde{\tau}) = 2.$$

The perturbation sequence $\epsilon^{(N)}$ is defined to be $\epsilon^{(N)}(1) = \epsilon^{(N)}(2) \equiv 0$ and $\epsilon^{(N)}(0) = 1/N$. A straightforward calculation shows, for $N \geq 2$,

$$\widetilde{\tau}^{(N)}(0, \omega_0) = 2, \quad \widetilde{\tau}^{(N)}(0, \omega_1) = 1.$$

Thus,

$$Y(0; \widetilde{\tau}^{(N)}) = 9/4 > 2 = Y(0; \widetilde{\tau}),$$

which is the claimed violation of stability.

The example paints a rather sceptical picture of the stability of the one-step-improvement. Indeed, the best we can now hope for, is (ia) there is a sequence $\bar{\tau}^{(N)}$ of stopping families such that

$$|\widetilde{\tau}^{(N)}(i) - \bar{\tau}^{(N)}(i)| \to 0 \quad P\text{-}a.s.$$

and, for all N, $\bar{\tau}^{(N)}$ is at least as good as $\widetilde{\tau}$, i.e.

$$Y(i; \bar{\tau}^{(N)}) \geq Y(i; \widetilde{\tau}).$$

(iia) The shortfall of $Y(i; \widetilde{\tau}^{(N)})$ below $Y(i; \widetilde{\tau})$ converges to zero P-a.s.

Note, however, that the convergence of the shortfall as in (iia) is the relevant question, not of the distance as in example (ii), page 35: the shortfall corresponds to a change for the worse of $\widetilde{\tau}^{(N)}$ compared to $\widetilde{\tau}$. As we are interested in an improvement it suffices to guarantee that such a change for the worse converges to zero. An additional improvement of $\widetilde{\tau}^{(N)}$ compared to $\widetilde{\tau}$ due to the error processes $\epsilon^{(N)}$ may be seen as a welcome side effect.

In the remainder of this section we sketch the proof of (ia) and (iia).

Theorem 3.1 *The one-step improvement is stable in the sense of (ia) and (iia).*

Remark 3.3 *It clearly suffices to prove (ia). Indeed,*

$$\left(Y(i; \widetilde{\tau}^{(N)}) - Y(i; \widetilde{\tau})\right)_- \leq \left(Y(i; \widetilde{\tau}^{(N)}) - Y(i; \bar{\tau}^{(N)})\right)_- + \left(Y(i; \bar{\tau}^{(N)}) - Y(i; \widetilde{\tau})\right)_- .$$

By (ia) the second term vanishes and the first converges to zero due to dominated convergence.

In order to construct an appropriate family $\bar{\tau}^{(N)}$ we first derive a criterion for a consistent stopping family $\bar{\tau}$ to be at least as good as $\widetilde{\tau}$. To this end define,

$$\widehat{\tau}(i) := \inf \left\{ j : \; i \leq j \leq k, \; Z(j) > \max_{j+1 \leq p \leq k} E^{\mathcal{F}_j} \left[Z(\tau(p)) \right] \right\}.$$

Obviously,

$$\widehat{\tau}(i) \geq \widetilde{\tau}(i).$$

Theorem 3.2 *Suppose τ, $\bar{\tau}$ are consistent stopping families and*

$$\widetilde{\tau}(i) \leq \bar{\tau}(i) \leq \widehat{\tau}(i). \tag{5}$$

Then,

$$Y(i; \bar{\tau}) \geq Y(i; \widetilde{\tau}) \geq \max \left\{ Z(i), \; \max_{p \geq i} E^{\mathcal{F}_i} \left[Z(\tau(p)) \right] \right\} \geq Y(i; \tau).$$

Proof. The last inequality is trivial, since $Y(i; \tau) = E^{\mathcal{F}_i} \left[Z(\tau(i)) \right]$. To prove the other inequalities we begin with a preliminary consideration. Define

$$\begin{aligned} \widetilde{Y}(i; \tau) &= \max_{p \geq i} E^{\mathcal{F}_i} \left[Z(\tau(p)) \right] \\ \widehat{Y}(i; \tau) &= \max_{p \geq i+1} E^{\mathcal{F}_i} \left[Z(\tau(p)) \right]. \end{aligned}$$

Then,

$$\widetilde{Y}(i;\tau) = \mathbf{1}_{\{\tau(i)>i\}}\widehat{Y}(i;\tau) + \mathbf{1}_{\{\tau(i)=i\}} \max\left\{\widehat{Y}(i;\tau),\ Z(i)\right\}, \tag{6}$$

since, by the consistency of τ,

$$\begin{aligned} E^{\mathcal{F}_i}\left[Z(\tau(i))\right] &= E^{\mathcal{F}_i}\left[\mathbf{1}_{\{\tau(i)=i\}}Z(i)\right] + E^{\mathcal{F}_i}\left[\mathbf{1}_{\{\tau(i)>i\}}Z(\tau(i+1))\right] \\ &= \mathbf{1}_{\{\tau(i)=i\}}Z(i) + \mathbf{1}_{\{\tau(i)>i\}}E^{\mathcal{F}_i}\left[Z(\tau(i+1))\right]. \end{aligned}$$

Step 1:

$$Y(i;\bar{\tau}) \geq \max\left\{Z(i),\ \max_{p\geq i} E^{\mathcal{F}_i}\left[Z(\tau(p))\right]\right\} \tag{7}$$

by backward induction over i. The induction base is obvious, since $\tau(k) = \bar{\tau}(k) = k$. Suppose now $0 \leq i \leq k-1$, and that the assertion is already proved for $i+1$. Note, $\{\bar{\tau}(i) = i\} \subset \{\widetilde{\tau}(i) = i\}$ by (5). Hence, we obtain on the set $\{\bar{\tau}(i) = i\}$,

$$Y(i;\bar{\tau}) = Z(i) \geq \widetilde{Y}(i;\tau).$$

However, on $\{\bar{\tau}(i) > i\}$ the induction hypothesis yields,

$$\begin{aligned} Y(i;\bar{\tau}) &= E^{\mathcal{F}_i}\left[Z(\bar{\tau}(i+1))\right] = E^{\mathcal{F}_i}\left[Y(i+1;\bar{\tau})\right] \geq E^{\mathcal{F}_i}\left[\widetilde{Y}(i+1;\tau)\right] \\ &= E^{\mathcal{F}_i}\left[\max_{i+1\leq p\leq k} E^{\mathcal{F}_{i+1}}\left[Z(\tau(p))\right]\right] \geq \max_{i+1\leq p\leq k} E^{\mathcal{F}_i}\left[Z(\tau(p))\right] \\ &= \widehat{Y}(i,\tau). \end{aligned}$$

Property (5) implies $\{\bar{\tau}(i) > i\} \subset \{\widehat{\tau}(i) > i\}$. Thus, on $\{\bar{\tau}(i) > i\}$,

$$\widehat{Y}(i,\tau) \geq Z(i)$$

and, by (6),

$$\widehat{Y}(i,\tau) = \widetilde{Y}(i,\tau) \quad \text{on } \{\bar{\tau}(i) > i\}.$$

This completes the proof of step 1. The second inequality now follows from (7) with the particular choice $\bar{\tau} = \widetilde{\tau}$.

Step 2: It remains to show that

$$Y(i;\bar{\tau}) \geq Y(i;\widetilde{\tau}).$$

For $i = k$ even equality holds. Suppose $0 \leq i \leq k-1$ and the inequality is proved for $i+1$. Then, on $\{\bar{\tau}(i) > i\} \cap \{\widetilde{\tau}(i) > i\}$,

$$Y(i,\bar{\tau}) = E^{\mathcal{F}_i}\left[Y(i+1,\bar{\tau})\right] \geq E^{\mathcal{F}_i}\left[Y(i+1,\widetilde{\tau})\right] = Y(i,\widetilde{\tau})$$

by induction hypothesis. On $\{\bar{\tau}(i) > i\} \cap \{\widetilde{\tau}(i) = i\}$

$$Y(i,\bar{\tau}) \geq Z(i) = Y(i,\widetilde{\tau})$$

by step 1. Finally, the set $\{\bar{\tau}(i) = i\} \cap \{\widetilde{\tau}(i) > i\}$ is evanescent by (5).

∎

Suppose, for the time being, the sequence $\widetilde{\tau}^{(N)}(i)$ converges P-a.s. to some stopping time $\bar{\tau}(i)$. Clearly, $\bar{\tau}$ is, as a limit of consistent stopping families, itself a consistent stopping family. It can be shown by backward induction over i, that $\bar{\tau}$ satisfies (5). Indeed, the basic idea is as follows. Assume $\bar{\tau}(i) = i$. Then, for $N \geq N_0(\omega)$ sufficiently large

$$\widetilde{\tau}^{(N)}(i) = i,$$

i.e.

$$Z(i) \geq \max_{i+1 \leq p \leq k} E^{\mathcal{F}_j}\left[Z(\tau(p))\right] + \epsilon^{(N)}(i).$$

We can now send N to infinity and obtain

$$Z(i) \geq \max_{i+1 \leq p \leq k} E^{\mathcal{F}_j}\left[Z(\tau(p))\right],$$

i.e.

$$\widetilde{\tau}(i) = i.$$

Thus, on $\{\bar{\tau}(i) = i\}$,

$$\widetilde{\tau}(i) \leq \bar{\tau}(i) \leq \widehat{\tau}(i).$$

A similar argument, making use of the induction hypothesis, yields the inequalities on $\{\bar{\tau}(i) > i\}$.

We can now define $\bar{\tau}^{(N)} = \bar{\tau}$ and (ia) is satisfied.

Unfortunately, example (page 35) shows that we may not expect $\widetilde{\tau}^{(N)}(i)$ to converge in general. Nonetheless, the previous considerations point to the right path. For ω such that $\widetilde{\tau}^{(M)}(i;\omega)$ converges, we define $\bar{\tau}^{(N)}(i;\omega)$ as this limit for all N. Otherwise, we define $\bar{\tau}^{(N)}(i;\omega) = i$, if and only if a subsequence of $\widetilde{\tau}^{(M)}(i;\omega)$ converges to i and $\widetilde{\tau}^{(N)}(i;\omega) = i$. The intuition is, that in the latter case we are free to choose the limit of any subsequence in order to obtain (5). So we choose $\bar{\tau}^{(N)}(i;\omega)$ as close as possible to $\widetilde{\tau}^{(N)}(i;\omega)$.

This reasoning can be formalized as follows. Define,

$$\bar{\tau}^{(N)}(k) = k$$

and

$$\begin{aligned}
\bar{\tau}^{(N)}(i) = i \quad &\iff \quad (\widetilde{\tau}^{(M)}(i) > i \text{ for only finitely many } M) \\
&\qquad \vee (\widetilde{\tau}^{(M)}(i) = i \text{ for infinitely many } M \text{ and } \widetilde{\tau}^{(N)}(i) = i) \\
\bar{\tau}^{(N)}(i) \neq i \quad &\implies \quad \bar{\tau}^{(N)}(i) = \bar{\tau}^{(N)}(i+1).
\end{aligned}$$

We have:

Lemma 3.3 *$\bar{\tau}^{(N)}$ satisfies (ia).*

The details of the proof can be found in Bender and Schoenmakers (2006), theorems 4.2 and 4.3. Stability of the algorithm, not only of one improvement step is also proven in this paper.

Remark 3.4 *Since $\bar{\tau}^{(N)}(i) \leq \widehat{\tau}(i)$, we obtain,*

$$\limsup_{N\to\infty} \widetilde{\tau}^{(N)}(i) \leq \widehat{\tau}(i).$$

On the other hand, the supermartingale property of the Snell envelope yields

$$\widehat{\tau}(i) \leq \widehat{\tau}^*(i) = \inf\left\{j: \ i \leq j \leq k, \ Z(j) > E^{\mathcal{F}_j}[Y^*(j+1)]\right\}.$$

As $\widehat{\tau}^*$ is the family of 'last optimal stopping times', we may conclude that the sub-optimality of $\widetilde{\tau}^{(N)}$ (for large N) basically stems from exercising to early.

4. NUMERICAL EXAMPLES

We now illustrate our algorithm with two examples: Bermudan basket-put and maximum-call options on 5 assets. We assume, that each asset is governed under the risk-neutral measure by the following SDE:

$$dS_i(t) = (r-\delta)S_i(t)dt + \sigma S_i(t)dW_i(t), \quad 1 \le i \le 5,$$

where $(W_1(t),\ldots,W_5(t))$ is a standard 5-dimensional Brownian motion. Suppose that an option can be exercised at $k+1$ dates $T_0,\ldots,T_k$, uniformly distributed between $T_0 = 0$ and $T_k = 3$(yr). The discounted price of the option is given by (1) with

$$Z(i) = e^{-rT_i}(K - \frac{S_1(T_i)+\cdots+S_5(T_i)}{5})^+ \quad \text{for the basket-put option and}$$
$$Z(i) = e^{-rT_i}(\max\{S_1(T_i),\ldots,S_5(T_i)\} - K)^+ \quad \text{for the maximum-call option.}$$

For our simulation, we take the following parameter values,

$$r = 0.05, \quad \sigma = 0.2, \quad S_1(0) = \ldots = S_5(0) = S_0, \quad K = 100,$$
$$\delta = 0 \text{ for basket-put option}, \quad \delta = 0.1 \text{ for maximum-call option.}$$

We consider options 'out-of-the-money', 'at-the-money' and 'in-the-money' at $t = 0$. For an initial stopping family $(\tau(i) :\ 0 \le i \le k)$, we construct the lower bound $Y(0;\tau)$, an improved lower bound $Y(0;\widetilde{\tau})$ with $\widetilde{\tau}$ given by (3), and the dual upper bound $Y_{up}(0;\tau)$ given by (4). A natural 'intuitively good' initial exercise rule is to exercise, when the cashflow is larger than the maximal value of all still-alive European options:

$$\tau(i) = \inf\{j \ge i : Z(j) \ge \max_{p \ge j+1} E^{\mathcal{F}_j} Z(p)\},$$

which is in fact a one-step improvement of the trivial exercise policy $\tau(i) \equiv i$. For our examples, however, a closed-form expression for still-alive Europeans $E^{\mathcal{F}_j} Z(p)$, $p > j$ does not exist. Fortunately, a good closed-form approximation is available for the basket-put option. For the maximum-call option we improve upon the exercise rule, suggested by Andersen (1999), Strategy 1. We will show that in all examples our method gives Bermudan prices with a relative accuracy better than 1%.

4.1. Bermudan basket-put

In this example we approximate still-alive European options by a moment-matching procedure. Let us define $f(T_j) := (S_1(T_j)+\cdots+S_5(T_j))/5$ for $0 \le j \le k$ and take j,p with $j \le p \le k$. First, we approximate $f(T_p)$ by

$$f_j(T_p) := f(T_j)\exp\left((r_j - \frac{1}{2}\sigma_j^2)(T_p - T_j) + \sigma_j(W(T_p) - W(T_j))\right),$$

where the parameters r_j and σ_j are taken in such a way that the first two moments of $f(T_p)$ and $f_j(T_p)$ are equal conditional $\mathcal{F}_j$:

$$r_j = r,$$
$$\sigma_j = \frac{1}{T_p - T_j} \ln \left(\frac{\sum_{m,n=1}^{5} S_m(T_j) S_n(T_j) \exp(1_{m=n}\sigma^2(T_p - T_j))}{(\sum_{m=1}^{5} S_m(T_j))^2} \right),$$

see, e.g., Brigo et al. (2004), Lord (2005). Then, we approximate $E^{\mathcal{F}_j} Z(p)$ by $E^{\mathcal{F}_j}[e^{-rT_p}(K - f_j(T_p))^+]$ using the Black-Scholes formula,

$$E^{\mathcal{F}_j}[e^{-rT_p}(K - f_j(T_p))^+] = e^{-rT_j} BS(f(T_j), r, \sigma_j, K, T_p - T_j),$$

and define the initial stopping family

$$\tau(i) := \{j \le i : \; Z(j) \ge e^{-rT_j} \max_{j+1 \le p \le k} BS(f(T_j), r, \sigma_j, K, T_p - T_j)\}, \quad 0 \le i \le k.$$

Note that the initial stopping family $(\tau(i) : \; 0 \le i \le k)$ leads already to a reasonable lower approximation $Y(0;\tau)$ of the Bermudan price (less then 5% relative). The gap between the improved lower bound $Y(0;\tilde{\tau})$ and dual upper bound $Y_{up}(0;\tau)$ does not exceeds 1% relative. See table 1, where we used 10^7 Monte Carlo trajectories for $Y(0;\tau)$ and 2000 trajectories (with 1000 nested trajectories) for $Y_{up}(0;\tau)$. To simulate $Y(0;\tilde{\tau})$ we use 10^5 outer and 500 inner trajectories. An obvious variance reduction is obtained by simulating $Y_{up}(0;\tau) - Y(0;\tau)$ and $Y(0;\tilde{\tau}) - Y(0;\tau)$ rather than $Y_{up}(0;\tau)$ and $Y(0;\tilde{\tau})$.

k	S_0	$Y(0;\tau)$ (SD)	$Y(0;\widetilde{\tau})$ (SD)	$Y_{up}(0;\tau)$ (SD)
3	90	10.000(0.000)	10.000(0.000)	10.000(0.002)
	100	2.156(0.001)	2.158(0.002)	2.162(0.001)
	110	0.537(0.001)	0.537(0.001)	0.538(0.001)
6	90	10.000(0.000)	10.000(0.000)	10.000(0.002)
	100	2.361(0.001)	2.395(0.004)	2.406(0.003)
	110	0.571(0.001)	0.578(0.002)	0.578(0.001)
9	90	10.000(0.000)	10.000(0.000)	10.001(0.002)
	100	2.387(0.001)	2.471(0.005)	2.490(0.006)
	110	0.579(0.001)	0.594(0.002)	0.596(0.002)

Table 1: Bermudan basket-put on 5 assets

4.2. Bermudan maximum-call

In contrast to the previous example, no good approximations are known for the still-alive maximum-call Europeans. For this example we take as initial stopping family strategy I of the Andersen method (see Andersen (1999)):

$$\tau(i) = \inf\{j \ge i : Z(j) \ge H_j\}.$$

The sequence of constants H_j is pre-computed using $5 \cdot 10^5$ simulations. Note that the gap between Andersen's lower bound $Y(0;\tau)$ and its dual upper bound $Y_{up}(0;\tau)$ varies from 2% to 4%, see columns 3 and 5 in table 2 (we use $5 \cdot 10^6$ Monte Carlo trajectories for $Y(0;\tau)$ and 500 Monte Carlo trajectories (with 1000 inner simulations for $Y_{up}(0;\tau) - Y(0;\tau)$). Further, we construct the improvement $Y(0;\widetilde{\tau})$ of Andersen's lower bound using 10^4 outer and 1000 inner simulations. The results are compared with the 90% confidence interval of Broadie and Glasserman (2004) computed by the stochastic mesh method, see table 2. We see that in almost all cases, $Y(0;\widetilde{\tau})$ and $Y_{up}(0;\tau)$ is within the 90% confidence interval, and that the gap between them does not exceed 1%.

Remark 4.1 *The cross-sectional least square algorithm by Longstaff and Schwartz (2001) yields results consistent with B-G: The lower bound reported in Longstaff and Schwartz (2001) for $d = 9$ and 19 basis functions are 16.657, 26.182, and 36.812, respectively. Slightly lower values are reported in Andersen and Broadie (2004) with 13 basis functions.*

k	S_0	$Y(0;\tau)$ (SD)	$Y(0;\widetilde{\tau})$ (SD)	$Y_{up}(0;\tau)$ (SD)	90% Confidence interval by BG
	90	15.702(0.008)	16.026(0.033)	15.986(0.021)	$[15.995,\ 16.016]$
3	100	24.716(0.009)	25.244(0.044)	25.333(0.031)	$[25.267,\ 25.302]$
	110	34.856(0.011)	35.695(0.056)	35.745(0.037)	$[35.679,\ 35.710]$
	90	16.064(0.007)	16.394(0.080)	16.462(0.054)	$[16.438,\ 16.505]$
6	100	25.171(0.009)	25.751(0.107)	25.978(0.066)	$[25.889,\ 25.948]$
	110	35.399(0.010)	36.329(0.131)	36.523(0.079)	$[36.466,\ 36.527]$
	90	16.202(0.007)	16.681(0.079)	16.734(0.063)	$[16.602,\ 16.710]$
9	100	25.343(0.009)	26.118(0.110)	26.333(0.083)	$[26.101,\ 26.211]$
	110	35.605(0.010)	36.652(0.134)	37.028(0.100)	$[36.719,\ 36.842]$

Table 2: Bermudan maximum-call on 5 assets

Concluding remarks

The iterative Monte Carlo procedures for pricing callable structures reviewed in this paper are quite generic as in principle it only requires a Monte Carlo simulation mechanism for an underlying Markovian system, for instance a Markovian system of SDEs. Moreover, by incorporating information obtained from another suboptimal method, for example Andersen's method (see Andersen (1999)) or the method of Longstaff and Schwartz (2001), we may improve upon this method to obtain our target results more efficiently.

The iterative procedures can be easily adapted to a large class of path-dependent exotic instruments where a call generates a sequence of cash-flows in the future. For these products one may construct 'virtual cash-flows' which are basically present values of future cash-flows specified in the contract. An important example is the (cancellable) snowball swap, an exotic interest rate product with growing popularity. In Bender et al. (2005) this product is treated in the context of a full-blown Libor market model (structured as in Schoenmakers (2005)). From this treatment it will be clear how to design Monte Carlo algorithms for related callable path-dependent Libor products.

References

L. Andersen. A simple approach to the pricing of Bermudan swaptions in a multifactor LIBOR market model. *J. Comput. Finance,* 3:5–32, 1999.

L. Andersen and M. Broadie. A primal-dual simulation algorithm for pricing multidimensional American options. *Manag. Science*, 50:1222–1234, 2004.

C. Bender and J. Schoenmakers. An iterative method for multiple stopping: Convergence and stability. *Adv. Appl. Prob.*, 2006. (forthcoming).

C. Bender, A. Kolodko, and J. Schoenmakers. Iterating snowballs and related path dependent callables in a multi-factor Libor model. *WIAS-Preprint No. 1061,* 2005. (submitted to *Risk* in modified form).

D. Brigo, F. Mercurio, F. Rapisarda, and R. Scotti. Approximated moment-matching dynamics for basket-options pricing. *Quantitative Finance*, 4:1–16, 2004.

M. Broadie and P. Glasserman. A stochastic mesh method for pricing high-dimensional American options. *J. Comput. Finance,* 7:35–72, 2004.

M. Broadie, P. Glasserman, and Z. Ha. Pricing American options by simulations using a stochastic mesh with optimized weights. In S. Uryasev, editor, *Probabilistic Constrained Optimization: Methodology and Applications*, pages 32–50. Kluwer: Norwell, Mass., 2000.

M. B. Haugh and L. Kogan. Pricing American options: A duality approach. *Operations Research*, 52:258–270, 2004.

R. Howard. *Dynamic programming and Markov processes*. MIT Press, Cambridge, Massachusetts, 1960.

A. Ibáñez and F. Zapatero. Valuation of American options through computation of the optimal exercise frontier. *J. Financial Quant. Anal.*, 39:253–276, 2004.

F. Jamshidian. LIBOR and swap market models and measures. *Finance and Stochastics,* 1:293–330, 1997.

A. Kolodko and J. Schoenmakers. Upper bound for Bermudan style derivatives. *Monte Carlo Meth. Appl.*, 10:331–343, 2004.

A. Kolodko and J. Schoenmakers. Iterative construction of the optimal Bermudan stopping time. *Finance and Stochastics*, 10:27–49, 2006.

F.A. Longstaff and R.S. Schwartz. Valuing American options by simulation: A simple least-square approach. *Review of Financial Studies*, 14:113–147, 2001.

R. Lord. A motivation for conditional moment matching. In M. Vanmaele, A. De Schepper, J. Dhaene, H. Reynaerts, W. Schoutens, and P. Van Goethem, editors, *Proceedings of the 3rd Actuarial and Financial Mathematics Day*, pages 69–94. Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten, 2005.

J. Neveu. *Discrete Parameter Martingales*. North-Holland, Amsterdam, 1975.

L.C.G. Rogers. Monte Carlo valuation of American options. *Math. Finance*, 12:271–286, 2002.

J. Schoenmakers. *Robust Libor modelling and pricing of derivative products*. Chapman & Hall – CRC Press, Boca Raton London New York Singapore, 2005.

EMPIRICAL ANALYSIS OF ANALYTIC APPROXIMATION APPROACHES FOR PRICING AND HEDGING SPREAD OPTIONS

Svetlana Borovkova, Ferry J. Permana and J.A.M. van der Weide

Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands
Email: `s.a.borovkova@ewi.tudelft.nl,f.j.permana@ewi.tudelft.nl,` `j.a.m.vanderWeide@ewi.tudelft.nl`

Abstract

In Borovkova et al. (2006), a new approach to valuation and hedging of basket options was developed, based on a generalized family of lognormal approximating distributions. This approach copes with possible negative values and negative skewness of a basket, and provides closed formulae for the option price and the greeks. This paper is devoted to a comparative simulation study of spread option pricing methods. We show that the Borovkova et al. (2006) approach performs well in terms of the option pricing and delta hedging, compared to other existing approximation approaches. Moreover, it is suited to baskets with several assets and with negative weights: a situation where other analytical approximation methods are not applicable. The analysis of the option's vegas shows that the price of a spread option is a decreasing function of the correlation and can decrease with the increase of individual volatilities – a seemingly paradoxical phenomenon of negative vegas.

1. INTRODUCTION

A basket option is an option whose payoff depends on the value of a basket, i.e., a portfolio of assets. A basket value is the weighted sum of individual asset prices, and, even when these prices have lognormal distribution, the basket value does not. This leads to difficulties in valuing and hedging basket options, similar to those arising in valuing Asian options.

Studies from many other areas of science (see e.g. Aitchison and Brown (1957), Mitchell (1968), Crow and Shimizu (1988), Limpert et al. (2001)) suggested to approximate the sum of lognormal random variables by the lognormal distribution, and have confirmed the high accuracy of such approximation. These results motivated the approach introduced in Borovkova et al. (2006), which is based on a generalized lognormal approximation of the basket value. Such an approximation makes it possible to apply the Black-Scholes formula to get a closed form expression for the value of a basket option.

There is a major obstacle to approximating a general basket distribution by the lognormal distribution: a basket with negative weights can have negative values and negative skewness, something that the regular lognormal distribution cannot approximate. To overcome this obstacle, in Borovkova et al. (2006) the basket distribution was approximated using the generalized family of lognormal distributions: regular, shifted, negative regular or negative shifted lognormal. We shall call this new approach the *Generalized Lognormal approach (GLN)*. These distributions approximate the basket distribution remarkably well and capture the features of general baskets, such as negative values and negative skewness, which cannot be captured using the regular lognormal distribution.

Using these approximating distributions, the Black-Scholes formula can be applied to obtain the option price and the greeks. The approach is easily implementable and it can deal with options on baskets with several assets and negative weights: a situation where most other existing analytical approximation approaches for pricing basket options cannot be applied.

A simplest basket with negative weights is a so-called *spread*, i.e., a difference between two asset prices. Several analytical approximation methods for pricing options on spreads were proposed: the Bachelier's method (applied by Shimko (1994)) and the Kirk's method (inspired by the classical paper of Margrabe (1978). A relatively new analytical approximation approach was proposed by Carmona and Durrleman (2003a). In Carmona and Durrleman (2003b) a possibility to extend their method to options on a linear combination of several assets was mentioned. However, nowadays it can only deal with spread options.

In Borovkova et al. (2006), the GLN approach was compared to the Bachelier and Kirk methods for spread options. A simulation study showed the GLN approach performs better than either of these methods in terms of both option pricing and delta-hedging. A simulation study in Carmona and Durrleman (2003b) also demonstrated the superiority of the Carmona method over the Kirk and Bachelier methods. In this paper we are concerned with comparing the performance of the GLN approach to the Carmona method.

In the next section, we review the GLN approach. Section 3 gives a short review of three other analytical approximation approaches. Empirical analysis of analytical approximation approaches is given in Section 4.

We consider baskets of futures on different (but related) commodities. Such basket options are very common in commodity markets. We also assume that the futures in the basket and the basket option mature at the same time. In practice, different commodity futures have different expiration schedules, and a typical basket option matures just before the earliest expiring futures or forward contract in the basket.

2. THE MODEL

Consider a basket of N futures, whose prices $F_i(t)$ follow correlated Geometric Brownian Motions. The basket value at time t is given by

$$B(t) = \sum_{i=1}^{N} a_i F_i(t),$$

where a_i is the weight corresponding to the asset i.

A general basket with negative weights can have negative values, and the basket distribution can be negatively skewed. Because of these features, the lognormal approximating distribution cannot be used directly.

In Borovkova et al. (2006), we propose to approximate the basket distribution using the generalized family of lognormal distributions: regular, shifted, negative regular or negative shifted lognormal. Recall that the probability density function (p.d.f.) of the regular log-normal distribution is given by

$$f(x) = \frac{1}{sx\sqrt{2\pi}} \exp\Big(-\frac{1}{2s^2}(\log x - m)^2\Big), x > 0. \tag{1}$$

If a random variable X has the (regular) log-normal distribution, then the random variable $Y = X + \tau$ has the shifted lognormal distribution and the random variable $Z = -X$ has the negative regular lognormal distribution. The combination of the shift and the reflection to the y-axis gives rise to the *negative shifted log-normal distribution*. m is the scale, s is the shape and τ is the shift parameter.

Recall that, under the risk-adjusted probability measure, the futures prices are martingales. Under the assumption of Geometric Brownian Motion dynamics of the futures prices (and hence the lognormality), the first three moments of the basket value on the maturity date T can be calculated. In terms of the first three moments, the skewness of basket is

$$\eta_{B(T)} = \frac{E\big[B(T) - EB(T)\big]^3}{s^3_{B(T)}}, \tag{2}$$

where $s_{B(T)} = \sqrt{EB^2(T) - (EB(T))^2}$ is the standard deviation of the basket value at time T.

For shifted and negative shifted lognormal distributions, we can derive the first three moments in terms of the parameters m, s, τ. The parameters of the appropriate approximating distribution are estimated by matching the first three moments of the basket with the first three moments of the appropriate log-normal distribution. This amounts to solving a nonlinear system of three equations with three unknowns (m, s and τ).

The skewness of basket distribution and the shift parameter play the key role in choosing the appropriate approximating distribution. Table 1 summarizes the choice of the approximating distribution for different parameter combinations.

Skewness	$\eta > 0$	$\eta > 0$	$\eta < 0$	$\eta < 0$
Shift parameter	$\tau \geq 0$	$\tau < 0$	$\tau \geq 0$	$\tau < 0$
Approximating distribution	regular	shifted	negative	neg.shifted

Table 1: Choice of the approximating distribution

Note that in the case $\eta > 0$, $\tau \geq 0$ our approach reduces to the Wakeman method (Turnbull and Wakeman (1991)), who approximate the distribution of a basket with positive weights by the lognormal distribution. If the basket distribution is assumed to be regular lognormal, then the basket option can be valued using the Black-Scholes formula (or in our case Black (1976)). For general baskets, the problem of pricing an option must be reduced to that simple case.

Let the basket 1 $(B_1(t))$ be (regular) log-normally distributed with parameters $m,\ s$. Furthermore, suppose that the basket 2 $(B_2(t))$ has the following relationship with the basket 1:

$$B_2(t) = B_1(t) + \tau$$

where τ is a constant. The distribution of basket 2 must be shifted log-normal with parameters $m,\ s,\ \tau$. On the maturity date T, the payoff of a call option on basket 2 is

$$\big(B_2(T) - X\big)^+ = \big((B_1(T) + \tau) - X\big)^+ = \big(B_1(T) - (X - \tau)\big)^+.$$

This is the payoff of a call option on the basket 1 with the same maturity date T and the strike price $(X - \tau)$, and such a call option can be valued by the Black's formula.

Using an analogous argument, we value a basket option using a negative lognormal distribution (see Borovkova et al. (2006)). Valuation of a basket option using a negative shifted lognormal distribution can be considered as a combination of valuation of a basket option with the shifted and negative regular lognormal distributions.

These arguments lead to the following closed form formulae for the price of a call option on a basket with the strike price X and time of maturity T. Everywhere $M_1(T)$ and $M_2(T)$ denote the first two moments of the basket on the maturity date T, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and $d_2 = d_1 - V$.

- Using the shifted log-normal approximation

$$c = e^{-rT}\big[(M_1(T) - \tau)\Phi(d_1) - (X - \tau)\Phi(d_2)\big] \tag{3}$$

$$\text{where} \quad d_1 = \frac{\log(M_1(T) - \tau) - \log(X - \tau) + \frac{1}{2}V^2}{V}$$

$$V = \sqrt{\log\left(\frac{M_2(T) - 2\tau M_1(T) + \tau^2}{(M_1(T) - \tau)^2}\right)}$$

- Using the negative log-normal approximation

$$c = e^{-rT}\big[-X\Phi(-d_2) + M_1(T)\Phi(-d_1)\big] \tag{4}$$

$$\text{where} \quad d_1 = \frac{\log(-M_1(T)) - \log(-X) + \frac{1}{2}V^2}{V}$$

$$V = \sqrt{\log\left(\frac{M_2(T)}{(M_1(T))^2}\right)}$$

- Using the negative shifted log-normal approximation

$$c = e^{-rT}\big[(-X - \tau)\Phi(-d_2) + (M_1(T) + \tau)\Phi(-d_1)\big] \tag{5}$$

$$\text{where} \quad d_1 = \frac{\log(-M_1(T) - \tau) - \log(-X - \tau) + \frac{1}{2}V^2}{V}$$

$$V = \sqrt{\log\left(\frac{M_2(T) + 2\tau M_1(T) + \tau^2}{(M_1(T) + \tau)^2}\right)}$$

In Borovkova et al. (2006), the closed formulae for the greeks are also derived.

3. OTHER ANALYTICAL APPROXIMATIONS

We shall compare the GLN approach to several analytical approximation approaches for spread option valuation. Note, however, that the GLN approach is not restricted to spread options, while other methods are.

Under the risk adjusted probability measure , the call price c on a spread $\big(F_2(t) - F_1(t)\big)$ at time 0, with time of maturity T and strike X, is given by :

$$c = e^{-rT} E\Bigg(F_2(T) - F_1(T) - X\Bigg)^+.$$

Several methods have been introduced to value spread options. The Bachelier method assumes the distribution of the spread can be approximated by the normal distribution. This allows for negative spread values, but not for a negative skewness. As a result, option prices obtained by the Bachelier method are often significantly different from the real option prices or those obtained by Monte Carlo simulation. The closed formula of a call price on a spread at time 0, with time of maturity T and strike X, is given by :

$$c^B = e^{-rT}\Bigg[F_2(0) - F_1(0) - X\Bigg]\Phi(d) + V^B\varphi(d)$$

where

$$\begin{aligned} d &= \frac{e^{-rT}\big(F_2(0) - F_1(0) - X\big)}{V^B} \\ V^B &= e^{-rT}\sqrt{F_2^2(0)\big(e^{\sigma_2^2 T} - 1\big) - 2F_1(0)F_2(0)\big(e^{\rho\sigma_1\sigma_2 T} - 1\big) + F_1^2(0)\big(e^{\sigma_1^2 T} - 1\big)} \end{aligned}$$

$\varphi(\cdot)$ is the probability density function of the standard normal distribution.

Another, more successful approximation method, is suggested by Kirk (1995), who replaced the difference of asset prices by the ratio, and adjusted the strike price. The closed formula of a call price on a spread at time 0, with time of maturity T and strike X, is given by :

$$c^K = e^{-rT}\Bigg[F_2(0)\Phi(d_1) - \big(F_1(0) + X\big)\Phi(d_2)\Bigg]$$

where

$$\begin{aligned} d_1 &= \frac{\log\left(\frac{F_2(0)}{F_1(0)+X}\right)}{V^K} + \frac{1}{2}V^K \\ d_2 &= \frac{\log\left(\frac{F_2(0)}{F_1(0)+X}\right)}{V^K} - \frac{1}{2}V^K \\ V^K &= e^{-rT}\sqrt{\sigma_2^2 T - 2\rho\sigma_2\sigma_2 T\bigg(\frac{F_1(0)}{F_1(0)+X}\bigg) + \sigma_2^2 T\bigg(\frac{F_1(0)}{F_1(0)+X}\bigg)^2}. \end{aligned}$$

A relatively new approach is proposed by Carmona and Durrleman (2003a). There, a good overview of spread options is given, and precise lower bounds are proposed to approximate spread option prices. The closed formula of a call price on a spread at time 0 with the time of maturity T and the strike X is given by :

$$c^C = e^{-rT}\left[F_2(0)\Phi\big(d^* + \sigma_2\sqrt{T}\cos(\theta^*+\phi)\big) - F_1(0)\Phi\big(d^* + \sigma_1\sqrt{T}\cos(\theta^*)\big) - X\Phi(d^*)\right]$$

where θ^* is a solution of following equation corresponding to the maximum :

$$\begin{aligned}
-\frac{1}{\sigma_2\sqrt{(T)}\cos(\theta+\phi)}\log\left[\frac{-\sigma_1 X\sin(\theta)}{F_2(0)\big(\sigma_2\sin(\theta+\phi)-\sigma_1\sin(\theta)\big)}\right] - \frac{1}{2}\sigma_2\sqrt{T}\cos(\theta+\phi)\\
= \frac{1}{\sigma_1\sqrt{(T)}\cos(\theta)}\log\left[\frac{-\sigma_2 X\sin(\theta+\phi)}{F_1(0)\big(\sigma_2\sin(\theta+\phi)-\sigma_1\sin(\theta)\big)}\right] - \frac{1}{2}\sigma_1\sqrt{T}\cos(\theta)
\end{aligned}$$

$$d^* = \frac{-1}{\sigma\sqrt{T}\cos(\theta^*-\psi)}\log\left[\frac{F_2(0)\sigma_2\sin(\theta^*+\phi)}{F_1(0)\sigma_1\sin(\theta^*)}\right] - \frac{1}{2}\big(\sigma_2\sqrt{T}\cos(\theta^*+\phi) - \sigma_2\sqrt{T}\cos(\theta^*)\big)$$

$$\phi = \arccos(\rho); \psi = \arccos(\frac{\sigma_1-\rho\sigma_2}{\sigma}); \sigma = \sqrt{\sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_1^2}.$$

However, there is a fundamental problem with the above equation: for certain (realistic) combinations of spread parameters, its solution does not exist, and hence, it is not clear how to apply the method proposed in Carmona and Durrleman (2003a). For such spreads the Carmona method cannot be applied directly.

4. SIMULATION STUDY

We apply our approach to a number of hypothetical spreads and baskets, such that all possible approximation distributions occur. We do not consider a regular normal approximating distribution since it reduces to the Wakeman method. The parameters of the test spreads are given in Table 2 and the call option prices on spreads are given in Table 3. For all spreads and baskets, the interest rate (r) is 3% per-annum and the time of maturity (T) is one year.

Our main motivation comes from basket options in energy markets, where typically assets have high volatilities and high correlations. However, we also apply our approach to a basket with a low correlation (spread 2) and low volatilities (spreads 2 and 3) to investigate the performance of our approach for low correlations and low volatilities as well. In all spreads and baskets, the options are almost in-the-money. The Monte Carlo simulation is repeated 1000 times for each spread (or basket), to obtain the means and standard errors of call prices (which are given in parenthesis in the last row of Table 3).

Simulation results in Table 3 show that both the Carmona and the GLN methods perform very well. The prices obtained by the Carmona method are slightly closer to those obtained by Monte Carlo. Note, however, that the GLN method also performs well for the basket consisting of three

	Spread 1	Spread 2	Spread 3	Basket 4	Basket 5
Futures price F_0	[100,110]	[120,100]	[200,50]	[95,90,105]	[100,90,95]
Volatility σ	[0.2,0.3]	[0.15,0.1]	[0.1;0.15]	[0.2,0.3,0.25]	[0.25,0.3,0.2]
Weights a	[-1,1]	[-1,1]	[-1,1]	[1,-0.8,-0.5]	[0.6,0.8,-1]
Correlation ρ	$\rho_{1,2}$=0.9	$\rho_{1,2}$=0.2	$\rho_{1,2}$=0.8	$\rho_{1,2}$=$\rho_{2,3}$=0.9 $\rho_{1,3}$=0.8	$\rho_{1,2}$=$\rho_{2,3}$=0.9 $\rho_{1,3}$=0.8
Strike price X	10	-20	-140	-30	35
Skewness η	$\eta > 0$	$\eta < 0$	$\eta < 0$	$\eta < 0$	$\eta > 0$
Shift parameter τ	$\tau < 0$	$\tau < 0$	$\tau > 0$	$\tau < 0$	$\tau < 0$

Table 2: Basket parameters

assets with negative weights (basket 4 and 5). The prices are almost everywhere within 95 % Monte Carlo confidence bounds, except for the spread 1. The deltas (w.r.t. the prices of futures 1 and 2) and vegas (w.r.t. σ_1, σ_2 and correlation ρ) for the spreads 1, 2 and 3 at time 0 are given in Table 4.

Method	Spread 1	Spread 2	Spread 3	Basket 4	Basket 5
GLN	6.7440 (shifted)	7.2643 (neg. shifted)	1.9576 (neg.regular)	7.7587 (neg.shifted)	9.0264 (shifted)
Carmona	6.7075	7.2560	1.9566	-	-
Kirk	6.7099	7.2350	1.5065	-	-
Bachelier	7.0004	7.3054	2.1214	-	-
Monte Carlo	6.7091 *(0.0126)*	7.2521 *(0.0098)*	1.9594 *(0.0045)*	7.7299 *(0.0095)*	9.0222 *(0.0151)*

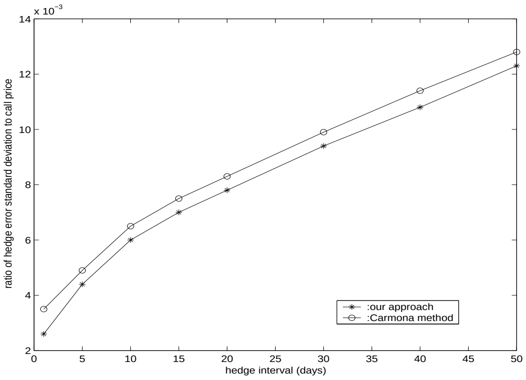
Table 3: Call option prices

Next, we investigate the performance of our method on the basis of delta-hedging the option, and compare it with Carmona method. We generate price paths of the basket assets from the time of writing the option until maturity, and on each hypothetical day we calculate the option's deltas with respect to each asset. We then re-adjust daily the hedging portfolio according to the deltas. We define the hedge error as the difference between the option price and the discounted hedge cost (i.e. the cost of maintaining the delta-hedged portfolio). If the hedging scheme works perfectly, the hedge cost would be exactly equal to the theoretical option price and the hedge error would be zero. In practice it is not zero due to the model error and discrete (e.g. daily) hedging. We expect the hedge error and its standard deviation to decrease when the hedge interval decreases, i.e. when hedging is done more frequently.

We investigate the delta-hedging performance for our approach on a spread option with parameters $F_0 = [100, 110]$, $\sigma = [0.1, 0.15]$, $a = [-1, 1]$, $\rho = 0.9$, $X = 10$, $r = 3\%$ per-annum and the time to maturity T one year. The spread distribution is approximated using a shifted lognormal. In

greeks	spread 1		spread 2		spread 3	
	GLN approach	Carmona	GLN approach	Carmona	GLN approach	Carmona
delta	-0.4573	-0.4599	-0.4639	-0.4548	-0.2200	-0.2200
	0.5149	0.4815	0.5274	0.4713	0.2226	0.2063
vega	-20.6582	-21.5671	39.6771	39.5018	55.4916	55.5976
	36.4075	36.6710	14.0768	13.1539	-8.3683	-8.4689
	-15.2418	-14.7080	-3.7561	-3.7017	-2.9813	-3.0064

Table 4: Deltas and Vegas



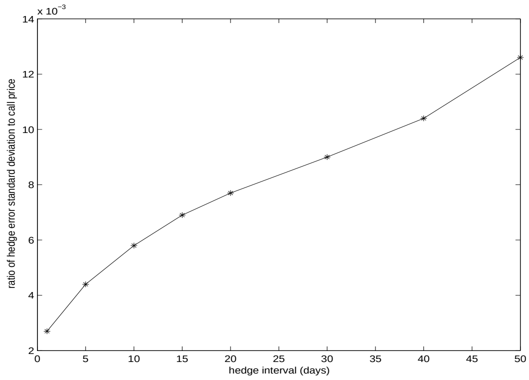


Figure 1: Left: Hedge error vs. hedge interval for a spread call option using GLM approach and Carmona method. Right: Hedge error vs. hedge interval for a basket call option consisting of 3 assets with negative weights, the GLN approach.

the left plot of figure 1 we show the ratio of the hedge error standard deviation to the call price vs. the hedge interval. The figure shows that for both the GLN approach and the Carmona method, this ratio (and so, the standard deviation of the hedge error) decreases together with the hedge interval, as we expect. In this example the mean hedge error is approximately the same (around 7%) for daily hedging for both methods. But the ratio obtained by the GLN approach is a slightly lower than that obtained by the Carmona method.

We also investigate the delta-hedging performance of the GLN approach for a basket consisting of 3 assets, one with positive and two with negative weights, with parameters $F_0 = [100, 90, 105]$, $\sigma = [0.2, 0.3, 0.25]$, $a = [1, -0.8, -0.5]$, $\rho_{1,2} = \rho_{2,3} = 0.9$, $\rho_{1,3} = 0.8$ $X = -30$, $r = 3\%$ per-annum and the time to maturity (T) is one year. We use a negative shifted lognormal distribution to approximate the basket distribution. The ratio of the hedge error standard deviation to the call price vs. the hedge interval shown is shown in the right of plot of Figure 1. The result is similar to that of the spread option in the previous example. For this basket, the mean of the hedge error is 4.8% for daily hedging.

Characteristic features of option vegas for spread options are different from the Black-Scholes model. Carmona and Durrleman (2003b) showed that the price of a spread option is a decreasing function of the correlation parameter. They also demonstrated on an example that the volatility vegas can be negative as well as positive. It means that the call price does not necessarily increase with increasing individual volatilities. The GLN approach applied to the spread option with parameters $F_0 = [110, 100]$, $a = [-1, 1]$, $r = 3\%$, $T = 1$ year, $\sigma = [0.3, 0.2]$ (*left figure*), $\rho = 0.9$

and $X = -10$ (*middle and right figures*) demonstrates the same phenomena, as shown in Figure 2. The observations are similar to those obtained by the Carmona method.

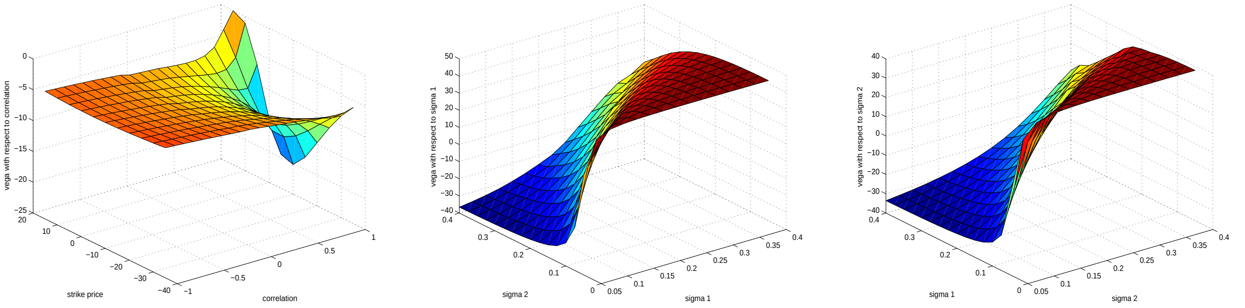


Figure 2: Left: Correlation vega vs. correlations and strike prices. Middle: Volatility vega w.r.t. σ_1 for different volatilities σ_1 and σ_2. Right: Volatility vega w.r.t. σ_2 for different volatilities σ_1 and σ_2.

5. CONCLUSION AND FUTURE WORK

Simulation studies on a number of hypothetical spreads showed that the Generalized Lognormal approach for valuing and hedging spread options, suggested by Borovkova et al. (2006) performs well, and its performance is comparable to the method proposed by Carmona and Durrleman (2003a). Moreover, the GLN approach has several advantages over the Carmona method, such as its applicability to baskets more general than two-assets spreads. A closed formula for the option price is derived by applying the Black-Scholes (or Black) formula, which is easy to implement and easy to be understand by practitioners, who are familiar with the Black-Scholes model. Application of the GLN approach to spread options confirms 'the negative vega' phenomenon, reported by Carmona and Durrleman (2003a).

In this paper we considered baskets of futures contracts. The GLN approach can be easily extended to baskets of physical commodities, as those considered by Carmona and Durrleman (2003a). The extension of the GLN approach to physical commodity baskets will be reported shortly.

An important feature of energy markets is that most delivery contracts are priced on the basis of an average price over a certain period. Hence, most energy derivatives (also basket and spread options) are Asian-style. So Asian basket options (that is, an Asian option on a basket of assets) need to be considered as well. Extension of our approach for valuation and hedging of Asian basket option is a topic of ongoing research.

Acknowledgement

The authors would like to thank Michel Dekking for many useful discussions.

References

J. Aitchison and J.A.C. Brown. *The log-normal distribution*. Cambridge University Press, Cambridge(UK), 1957.

F. Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3:167–179, 1976.

S. Borovkova, F.J. Permana, and J.A.M. v.d. Weide. A closed form approach to the valuation and hedging of basket and spread options. 2006. Submitted.

R. Carmona and V. Durrleman. Pricing and hedging spread options. *SIAM Rev.*, 45(4):627–685 (electronic), 2003a. ISSN 1095-7200.

R. Carmona and V. Durrleman. Pricing and hedging basket options in a log-normal model. Technical report, Department of Operations Research and Financial Engineering, Princeton University, 2003b.

E.L. Crow and K. Shimizu. *Log-normal Distributions:Theory and Application*. Dekker, New York, 1988.

E. Kirk. Correlation in the energy markets. *Managing energy price risk*, pages 71–78, 1995.

E. Limpert, W.A. Stahel, and M. Abbt. Lognormal distribution across the sciences: keys and clues. *Bioscience*, 51(5):341–352, 2001.

W. Margrabe. The value of an option to exchange one asset for another. *Journal of Finance*, 33 (1):177–86, 1978.

R.L. Mitchell. Permanence of the log-normal distribution. *Journal of the Optical Society of America*, 58(9):1267–1272, 1968.

D. Shimko. Options on futures spreads – hedging, speculation, and valuation. *Journal of Futures Markets*, 14:183–213, 1994.

S.M. Turnbull and L.M. Wakeman. A quick algorithm for pricing european average. *Journal of Financial and Quantitative Analysis*, 26(3):377–389, 1991.

ORNSTEIN-UHLENBECK MODELS FOR CREDIT RISK

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

[†]*European Commission, Joint Research Centre, I-21020 Ispra, Italy and K.U.Leuven, U.C.S., W. De Croylaan 54, B-3001 Leuven, Belgium.*
[§]*K.U.Leuven, U.C.S., W. De Croylaan 54, B-3001 Leuven, Belgium.*
Email: `Jessica.Cariboni@jrc.it`, `Wim.Schoutens@wis.kuleuven.ac.be`

Abstract

This work presents a reduced-form credit risk model driven by pure-jumps Ornstein-Uhlenbeck (OU) process. We analyse the case of the Gamma and Inverse Gaussian OU processes and show that the default probability can be expressed in closed-form through the characteristic function of the integrated OU process. The model is calibrated to a series of real-market credit default swap term structures. Results are compared with the well known cases of Poisson and CIR dynamics. We finally price a digital default put and show that models with pretty similar survival probabilities result in sometimes different option prices.

1. INTRODUCTION

Credit risk models are usually classified into two categories: structural models and reduced-form models. In structural models an event of default is defined in terms of boundary conditions on the asset value process. The first structural models date back to Merton (1974) and Black and Cox (1976) but a lot of modifications/extentions can be found in the literature (e.g. Leland (1994), Leland (1995), Madan et al. (1998), Cariboni and Schoutens (2004)). On the other hand, the reduced form approach models directly the default intensity and defines the time of default as the first jump-time of a counting process. The first example is given by the Jarrow and Turnbull (1995) model, who considered a constant default intensity. Subsequent generalizations allow for time-dependent or stochastic default intensities. In this latter case, the corresponding counting process is called a Cox-process. Duffie and Singleton (1999) developed a basic affine model, which allows for jumps in the hazard dynamics.

This work introduces a reduced-form model where the intensity of default follows a Ornstein-Uhlenbeck (OU) process. Under this assumption, we show that the survival probability of the obligor can be expressed in terms of the characteristic function of the integrated OU process. We consider the special cases of the Gamma-OU and Inverse Gaussian-OU (IG-OU) processes, where the characteristic function of the integrated process is available in closed-form. This allow to easily estimate the survival probability and price a broad class of credit derivatives.

We calibrate the model on a series of real Credit Default Swaps (CDS) term structures. For comparison purpose, we also calibrate intensity models based on the Poisson, inhomogeneous Poisson, and CIR dynamics. Once the models are calibrated, we price a digital default put and show that two models with pretty similar default probabilities and both almost perfectly calibrated on the market structures, can result in sometimes different option prices.

In the next section, we introduce the basic background on OU processes, concentrating on the Gamma and Inverse Gaussian OU processes. Section 3 presents our reduced-form OU default model. We then introduce CDS and link CDS spreads to the integrated OU process. The last part of section 3 presents the results of the calibration exercises and the pricing of the digital default put. The last section concludes.

2. ORNSTEIN-UHLENBECK PROCESSES

An OU process $y = \{y_t, t \geq 0\}$ (see e.g. Barndorff-Nielsen and Shephard (2001a), Barndorff-Nielsen and Shephard (2001b), Barndorff-Nielsen and Shephard (2003), Sato and Yamazato (1982)) is described by the following stochastic differential equation:

$$\mathrm{d}y_t = -\vartheta y_t \mathrm{d}t + \mathrm{d}z_{\vartheta t}, \quad y_0 > 0 \tag{1}$$

where ϑ is the arbitrary positive rate parameter and z_t is a subordinator, i.e. a Lévy process with no Brownian component, nonnegative drift and only positive increments. The process z_t is known as Background Driving Lévy Process (BDLP).

The process y_t is strictly stationary on the positive half-line, i.e. there exists a law D, called the stationary law, such that y_t will follow D for every t, if y_0 is chosen according to D (y_t is thus called D-OU process). In particular, given a one-dimensional distribution D there exists a D-OU process if and only if D is self-decomposable (for definition see Sato (1999)).

An important related process will be the so called integrated OU process (intOU) $Y = \{Y_t, t \geq 0\}$:

$$Y_t = \int_0^t y_s \mathrm{d}s.$$

One can show (see Barndorff-Nielsen and Shephard (2001a)) that for given y_0,

$$\begin{aligned}\log E[\exp(\mathrm{i}uY_t)|y_0] &= \vartheta \int_0^t k(u\vartheta^{-1}(1 - \exp(-\vartheta(t-s))))\mathrm{d}s \\ &\quad + \mathrm{i}uy_0\vartheta^{-1}(1 - \exp(-\vartheta t)),\end{aligned}$$

where $k(u) = k_z(u) = \log E[\exp(-uz_1)]$ is the cumulant function of z_1.

2.1. The Gamma-Ornstein-Uhlenbeck Process

The Gamma(a, b)-OU process has as BDLP a compound Poisson process:

$$z_t = \sum_{n=1}^{N_t} x_n$$

where $N = \{N_t, t \geq 0\}$ is a Poisson process with intensity a and $\{x_n, n = 1, 2, \ldots, N_t\}$ is a sequence of independent identically distributed Exp(b) variables. It turns out that the Gamma-OU process has a finite number of jumps in every compact time interval. Its stationary law is given by a Gamma(a, b) distribution:

$$f_{Gamma}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-xb), \quad x > 0,$$

which immediately explains the name.

For the Gamma-OU process, the characteristic function of the intOU process is given in closed-form by:

$$\begin{aligned} &\phi_{Gamma\text{-}OU}(u, t; \vartheta, a, b, y_0) = E[\exp(\mathrm{i}uY_t)|y_0] \\ &= \exp\left(\frac{\mathrm{i}uy_0}{\vartheta}(1 - \mathrm{e}^{-\vartheta t}) + \frac{\vartheta a}{\mathrm{i}u - \vartheta b}\left(b \log\left(\frac{b}{b - \mathrm{i}u\vartheta^{-1}(1 - \mathrm{e}^{-\vartheta t})}\right) - \mathrm{i}ut\right)\right). \end{aligned} \tag{2}$$

2.2. The Inverse Gaussian-Ornstein-Uhlenbeck Process

The Inverse Gaussian (IG(a, b)) density function is given by:

$$f_{IG}(x; a, b) = \frac{a}{\sqrt{2\pi}} \exp(ab) x^{-3/2} \exp(-(a^2 x^{-1} + b^2 x)/2), \quad x > 0.$$

This IG(a, b) is self-decomposable and hence an IG-OU process exists. The BDLP of a IG(a, b)-OU process is a sum of two independent Lévy processes $z = \{z_t = z_t^{(1)} + z_t^{(2)}, t \geq 0\}$. $z^{(1)}$ is an IG-Lévy process with parameters $a/2$ and b, while $z^{(2)}$ is of the form:

$$z_t^{(2)} = b^{-1} \sum_{n=1}^{N_t} v_n^2,$$

where $N = \{N_t, t \geq 0\}$ is a Poisson process with intensity parameter $ab/2$. $\{v_n, n = 1, 2, \ldots\}$ is a sequence of independent identically distributed random variables: each v_n follows a Normal$(0, 1)$ law independent from the Poisson process N. The IG-OU process jumps infinitely often in every interval via $z^{(1)}$.

The characteristic function of the integrated IG-OU process can also be given explicitly (see e.g. Nicolato and Venardos (2003)):

$$\begin{aligned} \phi_{IG\text{-}OU}(u, t; \vartheta, a, b, y_0) &= E[\exp(\mathrm{i}uY_t)|y_0] \\ &= \exp\left(\frac{\mathrm{i}uy_0}{\vartheta}(1 - \exp(-\vartheta t)) + \frac{2a\mathrm{i}u}{b\vartheta} A(u, t)\right), \end{aligned} \tag{3}$$

where

$$
\begin{aligned}
A(u,t) &= \frac{1-\sqrt{1+\kappa(1-\exp(-\vartheta t))}}{\kappa} \\
&\quad +\frac{1}{\sqrt{1+\kappa}}\left[\operatorname{arctanh}\left(\frac{\sqrt{1+\kappa(1-\exp(-\vartheta t))}}{\sqrt{1+\kappa}}\right)-\operatorname{arctanh}\left(\frac{1}{\sqrt{1+\kappa}}\right)\right], \\
\kappa &= -2b^{-2}\mathrm{i}u/\vartheta.
\end{aligned}
\tag{4}
$$

3. THE INTENSITY OU-MODEL

Reduced-form models assume an event of default to occur at the first jump of a counting process $M = \{M_t, t \geq 0\}$. The corresponding intensity of default $\lambda = \{\lambda_t, t \geq 0\}$ represents the instantaneous default probability:

$$
\lambda(t) = \lim_{h\to 0}\frac{P[\tau \in (t,t+h]|\tau > t]}{h}, \tag{5}
$$

where τ is the default time. The dynamics of the default intensity determine the credit quality of the corresponding asset.

We assume that the default intensity follows the Gamma-OU[1] or the IG-OU process introduced in the previous section. The dynamics are thus given by Equation (1):

$$
\mathrm{d}\lambda_t = -\vartheta\lambda_t \mathrm{d}t + \mathrm{d}z_{\vartheta t}, \quad \lambda_0 > 0.
$$

The time of default τ is defined as the first jump of M_t:

$$
\tau = \inf[t \in \mathbb{R}^+ \mid M_t > 0].
$$

The implied survival probability from 0 to t, $P(t)$, is given by:

$$
\begin{aligned}
P(t) &= P[M_t = 0] \\
&= P[\tau > t] \\
&= E\left[\exp\left(-\int_0^t \lambda_s \mathrm{d}s\right)\right] \\
&= E\left[\exp\left(-Y_t\right)\right] \\
&= \phi_{OU}(-\mathrm{i}, t; \vartheta, a, b, y_0),
\end{aligned}
\tag{6}
$$

where $\phi_{OU}(-\mathrm{i}, t; \vartheta, a, b, y_0)$ is the characteristic function of the intOU process evaluated at point $u = \mathrm{i}$. Equations (2) and (3) are used to evaluate the survival probability in the cases of the Gamma-OU and IG-OU processes.

[1]Note that the Gamma-OU case can be rephrased as a special case of the basic affine model introduced by Duffie and Singleton (1999).

3.1. Calibration of the model on CDS term structures

Credit Default Swaps (CDS) are derivatives that provide the buyer an insurance against the default of a company in exchange for (continuous) predetermined payments. The payments continue until the maturity of the contract, unless a default event occurs. In this case, the buyer delivers a bond on the underlying defaulting asset in exchange for its face value.

The price of a CDS with maturity T is given by the difference between the discounted spread and the loss payments:

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

where R is the recovery rate, r the risk free rate and $P(t)$ is the survival probability up to time t. The par spread c^* that makes this price equals to zero is:

$$\begin{aligned} c^* &= \frac{(1-R)\left(-\int_0^T \exp(-rs)\mathrm{d}P(s)\right)}{\int_0^T \exp(-rs)P(s)\mathrm{d}s} \\ &= \frac{(1-R)\left(1-\exp(-rT)P(T) - r\int_0^T \exp(-rs)P(s)\mathrm{d}s\right)}{\int_0^T \exp(-rs)P(s)\mathrm{d}s}. \end{aligned} \tag{7}$$

The closed-form expressions available for the survival probabilities in the cases of the Gamma and IG-OU dynamics allow to easily estimate c^*.

We calibrate the Gamma and IG-OU models to the CDS term structures ($T_1 = 1y$, $T_2 = 3y$, $T_3 = 5y$, $T_4 = 7y$, and $T_5 = 10y$ years) of the Itraxx Europe Index as of the 13th of December 2005. In this exercise we have set $r = 0.03$ and $R = 0.4$ for all the assets. In the calibrations we minimize the root mean square error ($rmse$):

$$rmse = \sqrt{\sum_{\text{CDS prices}} \frac{(\text{Market CDS price} - \text{Model CDS price})^2}{\text{number of CDS prices}}}.$$

The cpu time required to calibrate our OU-model to all the 125 CDS term structures is around one minute.

For comparison purposes, the capabilities of the OU model are tested by calibrating on the same term structures the following models:

1. the Homogeneous Poisson (HP) model (Jarrow and Turnbull (1995)), where the default intensity is constant;

2. the Inhomogeneous (INH) Poisson model with piecewise constant default intensity

$$\lambda_t = K_j, \quad T_{j-1} \le t < T_j, \quad j = 1, 2, \ldots, 5;$$

3. the Cox-Ingersoll-Ross (CIR) model (Cox et al. (1985)), where the default intensity is stochastic:

$$\mathrm{d}\lambda_t = \kappa(\eta - \lambda_t)\mathrm{d}t + \vartheta\sqrt{\lambda_t}\mathrm{d}W_t.$$

To compare the overall quality of the fits, we compute for each model the average absolute error as a percentage of the mean price (*ape*):

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

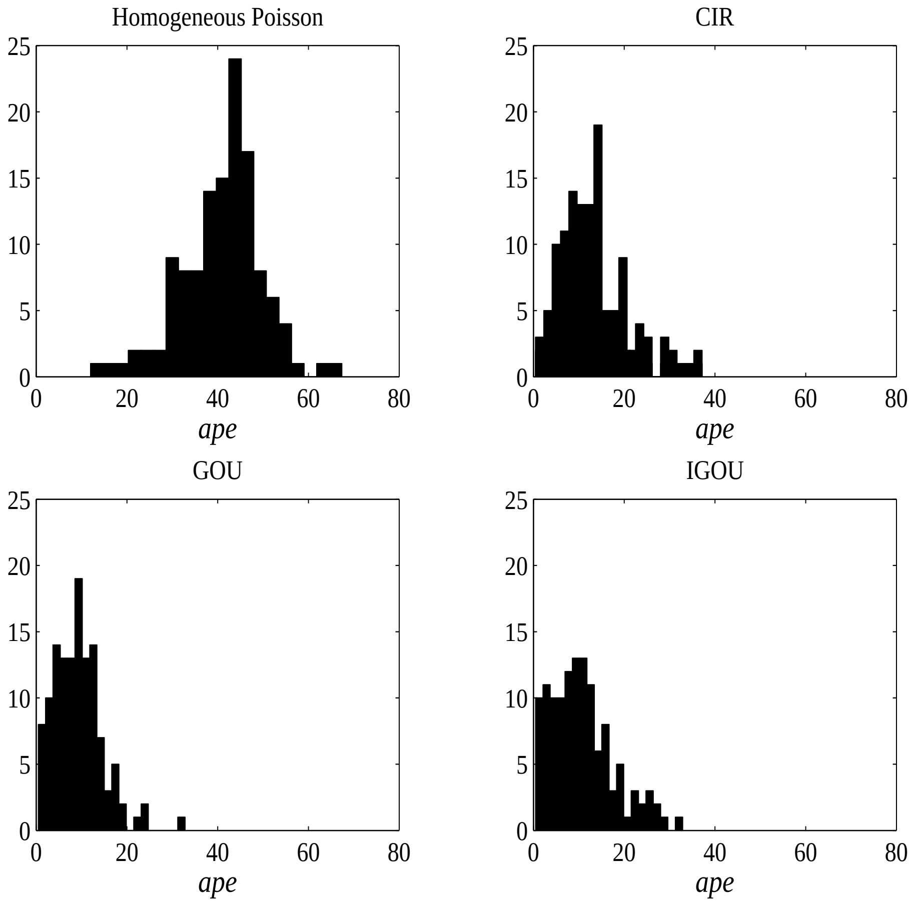


Figure 1: Distributions of the average absolute error as a percentage of the mean CDS price for the calibrated models.

Figure 1 plots the distributions of the *ape* for each model. Results for the inhomogeneous Poisson dynamics are not reported, since a perfect match is obtained between market and model prices (see below).

In the following, we concentrate on two companies, *ABN AMRO* and *TDC*. Similar results are obtained for all the other assets. Figures 2 and 3 plot the default probabilities (left plots) and the calibrated term structures (right plots). The figures highlight the failure of the HP model to match market data. The IHP model can match the market data but the behavior of the term structure is clearly unreliable, due to the piecewise constant assumption. The CIR, Gamma-OU and IG-OU models can all be nicely calibrated to market data. Figure 1 shows that, overall, the Gamma-OU dynamics outperforms the other two. Table 1 list market and models prices together with the values of the *rmse*.

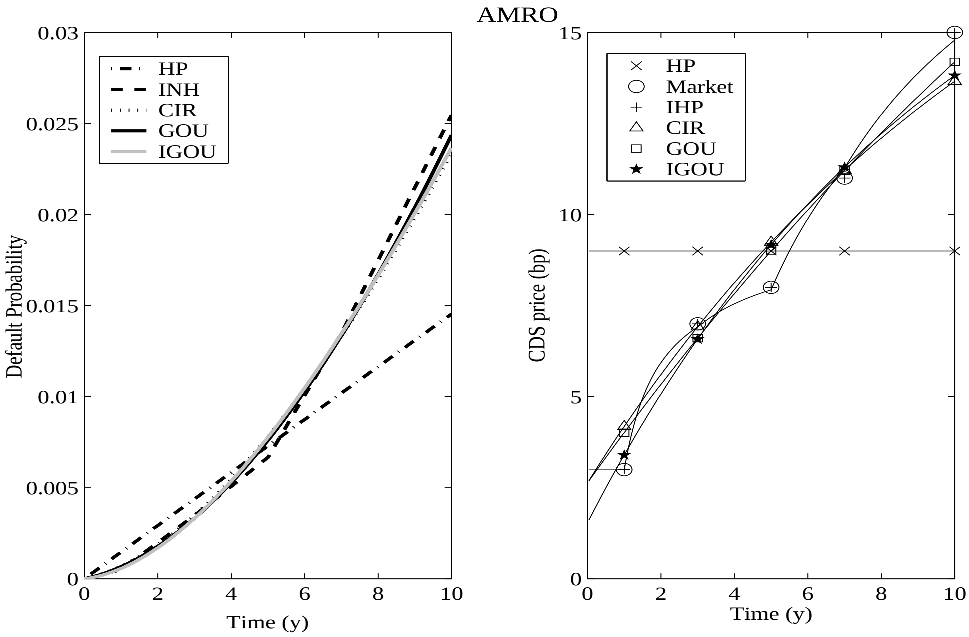


Figure 2: Estimated default probabilities and term structures for ABN AMRO Holding.

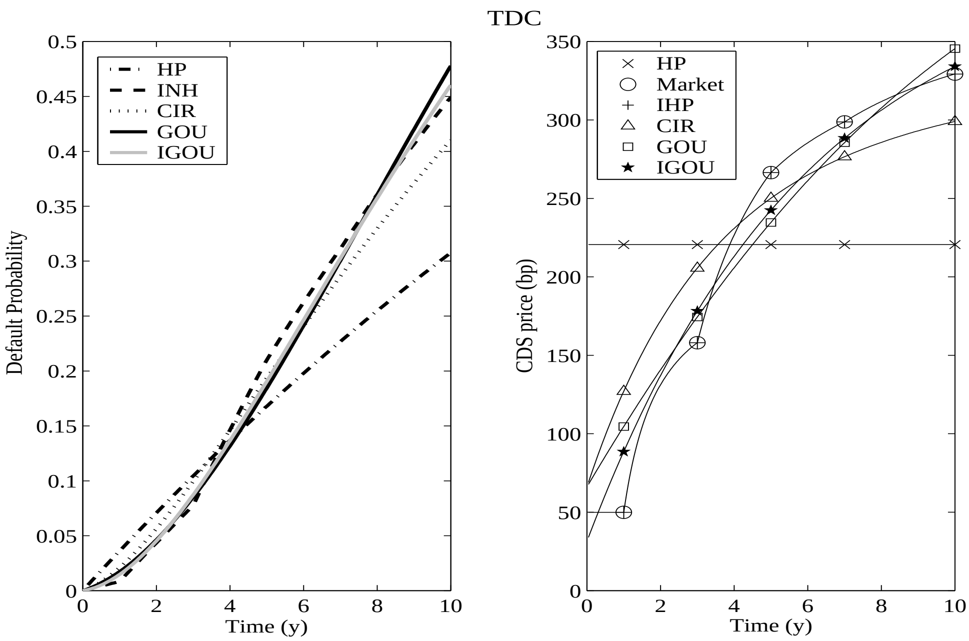


Figure 3: Estimated default probabilities and term structures for TDC.

Company		1y	3y	5y	7y	10y	*rmse*
ABN AMRO	Market	3	7	8	11	15	
	Model HP	9	9	9	9	9	8.95
	Model IHP	3	7	8	11	15	—
	Model CIR	4	7	9	11	14	2.10
	Model GOU	4	7	9	11	14	1.62
	Model IGOU	3	7	9	11	14	1.66
TDC	Market	50	158	266	299	329	
	Model HP	221	221	221	221	221	230.41
	Model IHP	50	158	266	299	329	—
	Model CIR	127	206	250	277	299	99.44
	Model GOU	105	174	235	286	345	68.616
	Model IGOU	89	178	242	288	334	51.10

Table 1: Results of the calibrations on CDS term structures (in bp).

3.2. Pricing of Digital Default Put and Model Risk

The calibrated models are finally used to price a Digital Default Put (DDP) with maturity T and payoff 1 at default. If default occurs at any time $\tau < T$, the owner of a DDP receives a unit payoff. The price of such an instrument at time $t < \tau$ is given by Schönbucher (2003):

$$D(t) = E\left[\int_t^T \lambda(s)\text{exp}\left(-\int_t^s (r(u)+\lambda(u))\mathrm{d}u\right)\mathrm{d}s\right]. \tag{8}$$

We estimate $D(t)$ using Monte Carlo simulation (sample size $N = 10\,000$). Figure 4 plots the prices for *AMRO* (left plot) and *TDC* (right plot). We concentrate here on the prices obtained with CIR, Gamma-OU and IG-OU dynamics, which best fit market data. Despite of the similar calibration results, the DDP prices for very low ($1y$) and very high ($10y$) times to maturity can be rather different. For intermediate time horizons some differences still exist but are less pronounced. If we focus on *TDC*, the maximum relative difference is obtained when comparing $1y$ prices (around 30% when comparing CIR and OU). Finally, although the calibrated CDS patterns are almost coincident, the same order of magnitude for the relative differences in DDP prices is obtained for *ABN AMRO* (around 12% when comparing IG-OU and CIR). This happens because of the path-dependence of the DDP price, which is not captured by the default probability behavior.

4. CONCLUSIONS

We have introduced a reduced-form credit risk model where the dynamics of the default intensity is described by a Gamma or Inverse Gaussian Ornstein-Uhlenbeck process. Under this assumption,

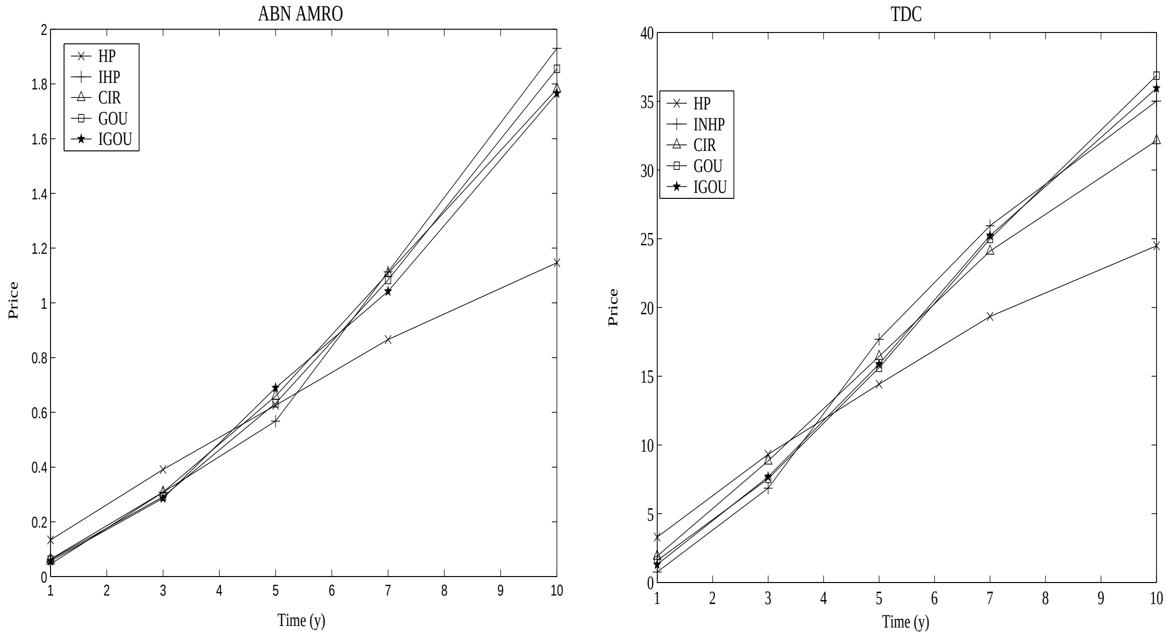


Figure 4: Price of the digital default put on ABN AMRO Holding (left plot) and TDC (right plot).

the survival probability has been expressed in closed-form using the characteristic function of the integrated process. We have shown that this allows to easily estimate the par spread of a CDS.

We have calibrated the model on the 125 CDS constituting the Itraxx Europe Index. The calibration of the model is quite fast: in one minute a standard pc station can calibrate the model to the complete set of 125 CDS. The Ornstein-Uhlenbeck model has been compared with the homogeneous and inhomogeneous Poisson models and with the Cox-Ingersoll-Ross dynamics. Results have shown that while homogeneous and inhomogeneous Poisson models fail in replicating real market structures, the CIR, Gamma-OU and IG-OU models can be nicely calibrated to market data. Generally, the Gamma-OU model outperforms the other two models in terms of mean squared difference between model and market prices.

After the calibrations, the models have been used to price a digital default put through Monte Carlo simulation. Despite of the similar calibration results, the option prices have sometimes resulted to be different (up to 30% of relative difference between e.g. the CIR and the Gamma-OU prices). This happens because of the path-dependence of the digital default put price, which is not captured by the default probability behavior.

References

O.E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B*, 63 (2):167–241, 2001a.

O.E. Barndorff-Nielsen and N. Shephard. Modelling by Lévy processes for financial econometrics. *In: O.E. Barndorff-Nielsen, T. Mikosch and S. Resnick (Eds.): Lévy Processes - Theory and Applications*, pages 283–318, 2001b.

O.E. Barndorff-Nielsen and N. Shephard. Integrated OU–processes and non-Gaussian OU–based stochastic volatility models. *Scandinavian Journal of Statistics*, 30(2):277–295, 2003.

F. Black and J. Cox. Valuing corporate securities: some effects on bond indenture provisions. *Journal of Finance*, 31(2):351–367, 1976.

J. Cariboni and W. Schoutens. Pricing credit default swaps under Lévy models. Technical Report 7, UCS, K.U.Leuven, Leuven, 2004.

J. Cox, J. Ingersoll, and S. Ross. A theory of the term structure of interest rates. *Econometrica*, 53 (2):385–408, 1985.

D. Duffie and K. Singleton. Modelling term structures of defaultable risky bonds. *Review of Financial Studies*, 12(4):687–720, 1999.

R. Jarrow and S. Turnbull. Pricing derivative on financial securities subject to credit risk. *Journal of Finance*, 50(1):53–86, 1995.

H. Leland. Corporate debt value, bond convenants, and optimal capital structure. *Journal of Finance*, 49(4):1213–1252, 1994.

H. Leland. A simple approach to valuing risky fixed and floating rate debt value. *Journal of Finance*, 50(3):789–819, 1995.

D.B. Madan, P. Carr, and E.C. Chang. The Variance–Gamma process and option pricing. *European Finance Review*, 2(1):79–105, 1998.

R. Merton. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29(2):449–470, 1974.

E. Nicolato and E. Venardos. Option pricing in stochastic volatility models of the Ornstein–Uhlenbeck type. *Mathematical Finance*, 13(4):445–466, 2003.

K. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge Studies in Advanced Mathematics 68. Cambridge University Press, Cambridge, 1999.

K. Sato and M. Yamazato. Stationary processes of Ornstein–Uhlenbeck type. In K. Itô and J.V. Prohorov, editors, *Probability Theory and Mathematical Statistics*, volume 1021 of *Lecture notes in mathematics*, pages 541–551. Birkhäuser, Boston, 1982.

P.J. Schönbucher. *Credit derivatives pricing models*. Wiley Finance, 2003.

STOP-LOSS PREMIUM BOUNDS ON MARKOV MARTINGALE PROCESSES

Cindy Courtois[†] and Michel Denuit[†§]

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

Abstract

In this paper, we deal with an incomplete market framework in a discrete time model. In actuarial science as well as in finance, the pricing of most products are based on some underlying stochastic process. Here, use is made of the theory of stochastic s-convex orderings and their respective extrema to find the extrema (minimum and maximum) for these underlying processes in incomplete markets. For example, the previous method can provide an analytic approach to the evaluation of aggregate claims models and the closely related stop-loss insurance.

As an application, we study the pricing problem of contingent claims of stop-loss type in the context of incomplete markets. As an illustration, the binomial and trinomial models are studied in detail. So, to calibrate the price of these products, our method leads to the computing of bounds within an incomplete market framework.

1. INTRODUCTION

Stochastic orderings are probabilistic tools to compare random variables or random vectors. Mathematically speaking, they are partial order relations defined on sets of probability distributions. Many papers have been devoted to the derivation of bounds in some stochastic order on a given random variable S. These bounds use some information about the random variable S, like moments, support, unimodality, etc. Relying on *extrema* with respect to some order relation, the actuary acts in a conservative way by basing his decisions on the least attractive risk that is consistent with the incomplete available information. The extrema correspond to the "worst" and the "best" risk. See, e.g., Denuit et al. (1999a) and the references therein.

In this paper, we will use the convex order, defined as follows: given two random variables S and T, S is said to be smaller than T in the convex order, denoted as $S \preceq_{\text{cx}} T$, if the inequality

$\mathbb{E}[\phi(S)] \leq \mathbb{E}[\phi(T)]$ holds for all the convex functions $\phi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist. The intuitive meaning of $S \preceq_{\text{cx}} T$ is that S is less variable than T. The multivariate version of $\preceq_{\text{cx}}$ is easily obtained by considering convex functions $\phi : \mathbb{R}^n \to \mathbb{R}$. However, this ordering does not allow for interesting applications (so that we will rather consider in this paper $\preceq_{\text{cx}}$-inequalities among linear combinations of the components of the random vectors to be compared, as explained below).

In this paper, we consider multiplicative discrete-time processes $\{X_n,\ n = 1, 2, \ldots\}$ obtained as follows. Starting from a sequence $\{Y_n,\ n = 1, 2, \ldots\}$ of positive independent random variables, we define recursively the X_n's as

$$X_{n+1} = X_n Y_{n+1},\ \ n = 1, 2, \ldots$$

with $X_1 = Y_1$. Such a process can be seen as a multiplicative random walk with relative increase Y_n at time n. It is widely used in finance to model the price of financial instruments (where X_n is the exponential of some process with independent increments). Our aim is to derive lower and upper bounds on the process $\{X_n,\ n = 1, 2, \ldots\}$ in the sense that any positive linear combination of the X_n's is bounded in the convex order by the corresponding linear combinations of the components of the extremal processes. This is similar to the works by Koshevoy and Mosler (1996 1997 1998) where orderings between random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ defined by $a_1X_1 + a_2X_2 + \cdots + a_nX_n \preceq_{\text{cx}} a_1Y_1 + a_2Y_2 + \cdots + a_nY_n$ for all constants $a_1, a_2, \ldots, a_n$ are studied.

The results derived in this paper are applied to discrete-time contingent claims pricing models. The underlying assets are assumed to follow a discrete-time process and trading only takes place at some prespecified dates. In this paper, we consider an incomplete market framework, so that the risk-neutral probability measure is not unique and we are in presence of a class of risk-neutral measures. The aim is thus to find the risk-neutral probability measures that imply the lower and upper bounds on the price of the claim and that are elements of the class of admissible prices. Examples within a trinomial model (i.e. a model where the change in the value of the stock between two trading times can attain three different values) are discussed.

The connection between the papers devoted to extremal distributions that appeared in the actuarial literature and financial pricing in incomplete markets is as follows. The class of risk-neutral probability measures is considered as a class of distributions with fixed support and first moment. Then, extremal elements are identified within the set of risk-neutral distributions, leading to bounds on the prices of contingent claims. This bridge between actuarial risk theory and financial mathematics seems to be promising.

The paper is organized as follows. The extremal processes are built in Section 2. Section 3 describes the application to financial pricing in the trinomial model. Numerical illustrations are provided there. The final Section 4 concludes.

2. EXTREMAL PROCESSES

2.1. Definitions

Let us denote as Y_i^- and Y_i^+ two positive random variables such that $Y_i^- \preceq_{\text{cx}} Y_i \preceq_{\text{cx}} Y_i^+$ holds for all i. Assume for instance that the support of Y_i is in $[a_i, b_i]$ and that $\mathbb{E}[Y_i] = \mu_i$. Then if we define the random variables Y_i^- and Y_i^+ as $Y_i^- = \mu_i$ almost surely, and

$$Y_i^+ = \begin{cases} a_i \text{ with probability } \dfrac{b_i - \mu_i}{b_i - a_i}, \\ b_i \text{ with probability } \dfrac{\mu_i - a_i}{b_i - a_i}, \end{cases}$$

we have $Y_i^- \preceq_{\text{cx}} Y_i \preceq_{\text{cx}} Y_i^+$. Other choices for the $\preceq_{\text{cx}}$-bounds are possible, according to the amount of information available about the Y_i's (support, moments, unimodality, ageing notions, etc.). See, e.g., Courtois and Denuit (2005) and the references therein.

All the random variables $Y_1, Y_2, \ldots, Y_1^-, Y_2^-, \ldots, Y_1^+, Y_2^+, \ldots$ are assumed to be independent. Starting from $X_1^- = Y_1^-$ and $X_1^+ = Y_1^+$, we define the extremal processes $\{X_n^-,\ n = 1, 2, \ldots\}$ and $\{X_n^+,\ n = 1, 2, \ldots\}$ by $X_i^- = X_{i-1}^- Y_i^-$ and $X_i^+ = X_{i-1}^+ Y_i^+$ for $i = 2, 3, \ldots$.

2.2. Convex ordered marginals

We expect that a convex ordering holds between X_i^-, X_i^+ and X_i. To prove that this is indeed the case, we will need the following useful lemma.

Lemma 2.1 *Let T_1, T_2, Z_1, Z_2 be independent and positive random variables such that $T_1 \preceq_{cx} T_2$ and $Z_1 \preceq_{cx} Z_2$. Then, $T_1 Z_1 \preceq_{cx} T_2 Z_2$ holds.*

Proof. Let ϕ be a convex function, and let us denote as F_{T_1}, F_{T_2}, F_{Z_1} and F_{Z_2} the distribution functions of T_1, T_2, Z_1 and Z_2, respectively. From

$$\begin{aligned} \mathbb{E}[\phi(T_1 Z_1)] &= \int_0^\infty \mathbb{E}[\phi(t Z_1)] dF_{T_1}(t) \\ &\leq \int_0^\infty \mathbb{E}[\phi(t Z_2)] dF_{T_1}(t) \text{ since } Z_1 \preceq_{\text{cx}} Z_2 \\ &= \int_0^\infty \mathbb{E}[\phi(T_1 z)] dF_{Z_2}(z) \\ &\leq \int_0^\infty \mathbb{E}[\phi(T_2 z)] dF_{Z_2}(z) \text{ since } T_1 \preceq_{\text{cx}} T_2 \\ &= \mathbb{E}[\phi(T_2 Z_2)], \end{aligned}$$

we conclude that the announced $\preceq_{\text{cx}}$-inequality indeed holds. ■

We are now ready to prove the next result that shows that the processes $\{X_n^-,\ n = 1, 2, \ldots\}$, $\{X_n,\ n = 1, 2, \ldots\}$ and $\{X_n^+,\ n = 1, 2, \ldots\}$ have indeed $\preceq_{\text{cx}}$-ordered univariate marginals.

Proposition 2.2 *The stochastic inequalities $X_i^- \preceq_{cx} X_i \preceq_{cx} X_i^+$ hold for all i.*

Proof. Let us prove the announced result using an iterative argument. The result is obviously true for $i = 1$, since it reduces to $Y_1^- \preceq_{\text{cx}} Y_1 \preceq_{\text{cx}} Y_1^+$. Now, assume that the result holds for

$i = 1, 2, \ldots, n$ and let us prove it for $n+1$. Let us apply Lemma 2.1 in our setting. Taking $T_1 = T_2 = X_n^-$ and $Z_1 = Y_{n+1}^-$, $Z_2 = Y_{n+1}$, we get

$$X_n^- Y_{n+1}^- = X_{n+1}^- \preceq_{\text{cx}} X_n^- Y_{n+1}.$$

Now, taking $T_1 = T_2 = Y_{n+1}$ and $Z_1 = X_n^-$, $Z_2 = X_n$, we have

$$X_n^- Y_{n+1} \preceq_{\text{cx}} X_n Y_{n+1} = X_{n+1}.$$

We then conclude that $X_{n+1}^- \preceq_{\text{cx}} X_{n+1}$ by transitivity. The proof of $X_{n+1} \preceq_{\text{cx}} X_{n+1}^+$ follows along the same lines. ■

2.3. Convex ordered linear combinations

Let us now prove that any positive linear combination of the X_i's is bounded from below and from above in the $\preceq_{\text{cx}}$-sense by the same combination of the X_i^-'s and of the X_i^+'s.

Proposition 2.3 *Whatever the positive constants $\alpha_1, \ldots, \alpha_n$, the stochastic inequalities*

$$\sum_{j=1}^{n} \alpha_j X_{i_j}^- \preceq_{cx} \sum_{j=1}^{n} \alpha_j X_{i_j} \preceq_{cx} \sum_{j=1}^{n} \alpha_j X_{i_j}^+$$

hold for any $i_1 < i_2 < \cdots < i_n$ and integer n.

Proof. We only prove the stochastic inequality $\sum_{j=1}^n \alpha_j X_{i_j}^- \preceq_{\text{cx}} \sum_{j=1}^n \alpha_j X_{i_j}$; the reasoning to establish $\sum_{j=1}^n \alpha_j X_{i_j} \preceq_{\text{cx}} \sum_{j=1}^n \alpha_j X_{i_j}^+$ is similar. The result is obviously true for $n = 1$. Let us first establish the result for $n = 2$. To this end, let us write

$$\begin{aligned} \alpha_1 X_{i_1}^- + \alpha_2 X_{i_2}^- &= X_{i_1}^- \big(\alpha_1 + \alpha_2 Y_{i_1+1}^- \ldots Y_{i_2}^-\big) \\ \alpha_1 X_{i_1} + \alpha_2 X_{i_2} &= X_{i_1} \big(\alpha_1 + \alpha_2 Y_{i_1+1} \ldots Y_{i_2}\big). \end{aligned}$$

Since $Y_{i_1+1}^- \ldots Y_{i_2}^- \preceq_{\text{cx}} Y_{i_1+1} \ldots Y_{i_2}$ and since $\preceq_{\text{cx}}$ is closed under changes of scale and origin, Lemma 2.1 then gives $\alpha_1 X_{i_1}^- + \alpha_2 X_{i_2}^- \preceq_{\text{cx}} \alpha_1 X_{i_1} + \alpha_2 X_{i_2}$, as announced. Now, let us assume that the result holds for n and let us establish it for $n+1$. First, note that

$$\alpha_1 X_{i_1}^- + \cdots + \alpha_{n+1} X_{i_{n+1}}^- = \alpha_1 X_{i_1}^- + \cdots + X_{i_n}^- \big(\alpha_n + \alpha_{n+1} Y_{i_n+1}^- \ldots Y_{i_{n+1}}^-\big).$$

The recurrence relation ensures that, given $Y_{i_n+1}^- \ldots Y_{i_{n+1}}^- = t$, the stochastic inequality

$$\alpha_1 X_{i_1}^- + \cdots + X_{i_n}^- \big(\alpha_n + \alpha_{n+1} t\big) \preceq_{\text{cx}} \alpha_1 X_{i_1} + \cdots + X_{i_n} \big(\alpha_n + \alpha_{n+1} t\big)$$

holds true. Since $Y_{i_n+1}^- \ldots Y_{i_{n+1}}^-$ is independent from both $X_{i_1}^-, \ldots, X_{i_n}^-$ and $X_{i_1}, \ldots, X_{i_n}$, the $\preceq_{\text{cx}}$-inequality also holds unconditionally, so that we get

$$\begin{aligned} \alpha_1 X_{i_1}^- + \cdots + \alpha_{n+1} X_{i_{n+1}}^- &\preceq_{\text{cx}} \alpha_1 X_{i_1} + \cdots + X_{i_n} \big(\alpha_n + \alpha_{n+1} Y_{i_n+1}^- \ldots Y_{i_{n+1}}^-\big) \\ &\preceq_{\text{cx}} \alpha_1 X_{i_1} + \cdots + X_{i_n} \big(\alpha_n + \alpha_{n+1} Y_{i_n+1} \ldots Y_{i_{n+1}}\big) \end{aligned}$$

since $Y_{i_n+1}^- \ldots Y_{i_{n+1}}^- \preceq_{\text{cx}} Y_{i_n+1} \ldots Y_{i_{n+1}}$. This ends the proof. ■

3. APPLICATIONS TO THE TRINOMIAL MODEL FOR STOCK PRICES

3.1. Description of the model

In the trinomial asset pricing model, we begin with an initial stock price $S_0 = 1$. There are three possible numbers, d, 1 and u, with $0 < d < 1 < u$, such that at the next period, the stock price will be either dS_0, S_0 or uS_0. Typically, we take d and u to satisfy $0 < d < 1 < u$, so change of the stock price from S_0 to dS_0 represents a *downward* movement, and change of the stock price from S_0 to uS_0 represents an *upward* movement. Therefore, at each time step, the stock price either goes up by a factor u or down by a factor d or does not move.

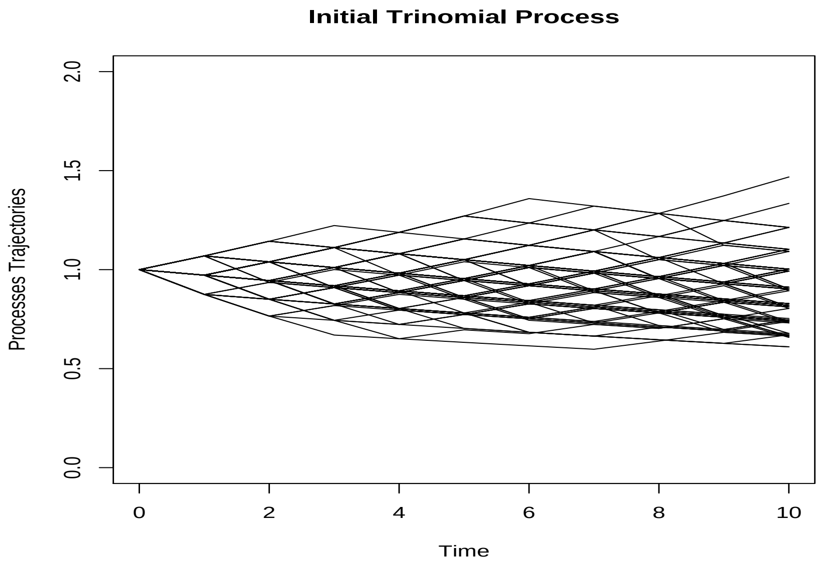


Figure 1: 100 *trajectories of the trinomial process.*

Let $\{S_n,\ n = 0, 1, \ldots\}$ be the stock price process and r be the risk-free interest rate. We also assume that $d < 1$ and $1 + r < u$ (no arbitrage opportunities). This process falls into the scope of this paper since S_{n+1} can be obtained as the product of the previous stock price S_n times J_{n+1}, where the J_n's are independent and identically distributed random variables, taking the values d, 1 or u. Figure 1 describes 100 typical trajectories of the trinomial process with $u = 1.1$, $d = 0.9$. The physical probabilities associated with the downward (d), stationary (1) and upward (u) movements are respectively equal to 10%, 51.26% and 38.74% as in Hull (2002).

The financial pricing of contingent claims is not made under the physical (or historical) probability distribution, but well under the risk-neutral one. Recall that a *risk-neutral probability measure* is a measure that agrees with the *physical probability measure* about which price paths have zero probability, and under which the discounted prices of all primary assets are martingales. The condition for the model to be free of arbitrage opportunities is the existence of a risk-neutral probability measure and the price is then obtained by taking the expectation of the discounted payoff under such a measure.

If there exist claims that are not attainable, then the market is said to be *incomplete*. In this case there are infinitely many risk-neutral measures. The trinomial model is known to be incomplete. The space $\mathcal{P}$ of all risk-neutral probability measures is taken such that under all risk-neutral probability measures $\widetilde{\mathbb{P}}$, the discounted stock price process $\{\frac{S_n}{(1+r)^n},\ n=0,1,\ldots\}$ is a martingale with respect to the natural filtration, i.e.

$$\mathbb{E}_{\widetilde{\mathbb{P}}}\left[\frac{S_{n+1}}{(1+r)^{n+1}}\middle|S_0,S_1,\ldots,S_n\right]=\frac{S_n}{(1+r)^n}\quad\text{for any }\widetilde{\mathbb{P}}\in\mathcal{P}.$$

3.2. The set $\mathcal{P}$ of risk-neutral probability measures

Let us denote as X_n the discounted stock price, that is, $X_n=\frac{S_n}{(1+r)^n}$, starting from $X_0=S_0$. The process $\{X_n,\ n=1,2,\ldots\}$ admits the representation $X_n=X_{n-1}Y_n$ with

$$(1+r)Y_n=\begin{cases} d \text{ with probability } \widetilde{p}_2, \\ 1 \text{ with probability } \widetilde{p}_1, \\ u \text{ with probability } \widetilde{p}_3. \end{cases}$$

By convention, $X_1=Y_1$.

Within the trinomial model, every risk-neutral probability measure $\widetilde{\mathbb{P}}$ corresponds to a triplet $(\widetilde{p}_1,\widetilde{p}_2,\widetilde{p}_3)$ of positive real numbers satisfying $\widetilde{p}_1+\widetilde{p}_2+\widetilde{p}_3=1$, where the risk-neutral probabilities $\widetilde{p}_1$, $\widetilde{p}_2$ and $\widetilde{p}_3$ are respectively associated with a stationary (1), downward (d) and upward (u) movement of the stock price process. The class $\mathcal{P}$ of risk-neutral probability measures can then be identified with the set of admissible triplets.

All the risk-neutral probability measures, henceforth denoted as $\widetilde{\mathbb{P}}=(\widetilde{p}_1,\widetilde{p}_2,\widetilde{p}_3)$, must be equivalent to the historical measure (in the sense that the set of events that have probability 0 under $\widetilde{\mathbb{P}}$ is the same as the set of events that have probability 0 under the physical measure $\mathbb{P}$) and such that $\mathbb{E}_{\widetilde{\mathbb{P}}}[X_{n+1}|X_n]=X_n$ for all n. So $(\widetilde{p}_1,\widetilde{p}_2,\widetilde{p}_3)$ have to verify the following system

$$\begin{cases} \widetilde{p}_1+\widetilde{p}_2+\widetilde{p}_3=1 \\ \widetilde{p}_2\cdot d+\widetilde{p}_1\cdot 1+\widetilde{p}_3\cdot u=1+r \end{cases}$$

with $0<\widetilde{p}_i<1$ ($i=1,2,3$). This is equivalent to say that all risk-neutral probability measures $(\widetilde{p}_1,\widetilde{p}_2,\widetilde{p}_3)$ must be such that $\widetilde{p}_1=1-\frac{r}{u-1}-\widetilde{p}_2\frac{u-d}{u-1}$ and $\widetilde{p}_3=\frac{r}{u-1}+\widetilde{p}_2\frac{1-d}{u-1}$ with $0<\widetilde{p}_2<\frac{u-(1+r)}{u-d}$.

3.3. Extremal price processes

Denuit and Lefèvre (1997) and Denuit et al. (1999b) derived $\preceq_{\text{cx}}$-bounds on random variables valued in $\{0,1,\ldots,n\}$. These extremal distributions can be generalized to the case of random variables valued in an arbitrary set $\mathcal{E}_n=\{e_0,\ldots,e_n\}$, with $e_0<e_1<\cdots<e_n$, in the spirit of Denuit et al. (1999c). Specifically, consider a random variable S valued in $\mathcal{E}_n$ with mean μ. Defining

$$S_{\min}^{\text{disc}}=\begin{cases} e_k \text{ with probability } \dfrac{e_{k+1}-\mu}{e_{k+1}-e_k}, \\[2ex] e_{k+1} \text{ with probability } \dfrac{\mu-e_k}{e_{k+1}-e_k}, \end{cases}$$

where $e_k \in \mathcal{E}_{n-1}$ is such that $e_k < \mu \leq e_{k+1}$, and

$$S_{\max}^{\text{disc}} = \begin{cases} e_0 \text{ with probability } \dfrac{e_n - \mu}{e_n - e_0}, \\ e_n \text{ with probability } \dfrac{\mu - e_0}{e_n - e_0}, \end{cases}$$

we have $S_{\min}^{\text{disc}} \preceq_{\text{cx}} S \preceq_{\text{cx}} S_{\max}^{\text{disc}}$. Knowing that $\mathbb{E}_{\widetilde{\mathbb{P}}}[Y_n] = 1$ for all n, we see easily that the random variables Y_n^- and Y_n^+ such that the stochastic inequalities $Y_n^- \preceq_{\text{cx}} Y_n \preceq_{\text{cx}} Y_n^+$ hold true are defined by

$$(1+r)Y_n^- = \begin{cases} 1 \text{ with probability } \dfrac{u-(1+r)}{u-1}, \\ u \text{ with probability } \dfrac{r}{u-1}, \end{cases}$$

and

$$(1+r)Y_n^+ = \begin{cases} d \text{ with probability } \dfrac{u-(1+r)}{u-d}, \\ u \text{ with probability } \dfrac{(1+r)-d}{u-d}. \end{cases}$$

The processes $\{X_n^-,\ n = 1, 2, \ldots\}$ and $\{X_n^+,\ n = 1, 2, \ldots\}$ are then defined by $X_n^- = X_{n-1}^- Y_n^-$ and $X_n^+ = X_{n-1}^+ Y_n^+$, starting from $X_1^- = Y_1^-$ and $X_1^+ = Y_1^+$.

The stochastic processes $\{X_n^-,\ n = 1, 2, \ldots\}$ and $\{X_n^+,\ n = 1, 2, \ldots\}$ are trinomial models with probabilities associated to $(1, d, u)$ being respectively $\left(\frac{u-(1+r)}{u-1}, 0, \frac{r}{u-1}\right)$ $\left(0, \frac{u-(1+r)}{u-d}, \frac{(1+r)-d}{u-d}\right)$. These two sets of probabilities do not correspond to risk neutral measures (since the support is not the physical one). The two extremal processes are obtained by letting the probability associated to d (i.e. $\widetilde{p}_2$) converging to its minimal and maximal possible values (i.e. 0 and $\frac{u-(1+r)}{u-d}$). Figure 2 describes 100 trajectories of the minimal and the maximal processes with $u = 1.1$, $d = 0.9$.

3.4. Numerical results

3.4.1. European call option

The owner of a European call option has the right to buy a stock for K (strike price) at a certain future time N. We denote by S_0 the current value of the stock price and we make the assumption that the considered stock price follows a trinomial model with N periods of time. Considering $\widetilde{\mathbb{P}}$ in the set $\mathcal{P}$ of risk-neutral probability measures, a possible price of this European call is given by $\frac{1}{(1+r)^N}\mathbb{E}_{\widetilde{\mathbb{P}}}[(S_N - K)_+]$. Every possible price satisfies

$$\frac{1}{(1+r)^N}\mathbb{E}_{\widetilde{\mathbb{P}}}[(S_N^- - K)_+] \leq \frac{1}{(1+r)^N}\mathbb{E}_{\widetilde{\mathbb{P}}}[(S_N - K)_+] \leq \frac{1}{(1+r)^N}\mathbb{E}_{\widetilde{\mathbb{P}}}[(S_N^+ - K)_+],$$

where

$$\mathbb{E}_{\widetilde{\mathbb{P}}}[(S_N^- - K)_+] = \sum_{i=0}^{N} \binom{N}{i} \left(\frac{r}{u-1}\right)^i \left(\frac{u-(1+r)}{u-1}\right)^{N-i} (S_0 u^i - K)_+$$

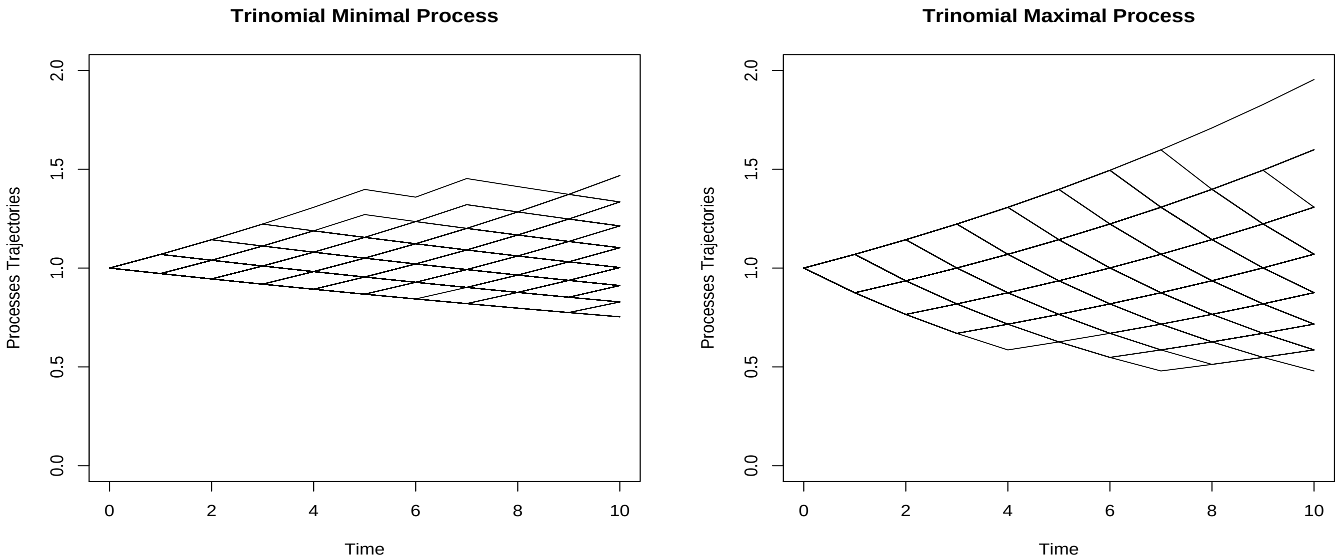


Figure 2: 100 *trajectories of the extremal trinomial process.*

and

$$\mathbb{E}_{\widetilde{\mathbb{P}}}[(S_N^+ - K)_+] = \sum_{i=0}^{N} \binom{N}{i} \left(\frac{(1+r)-d}{u-d}\right)^i \left(\frac{u-(1+r)}{u-d}\right)^{N-i} (S_0 u^i d^{N-i} - K)_+.$$

Table 1 displays the bounds obtained on the call price for different maturities and strike prices. As in Hull (2002), we consider $u = 1.1$ and $d = 0.9$. The annual risk-free rate is 12%. A period of time corresponds to 3 months. The range of possible values for the price of the call option is not too large.

N	K	Minimum	Maximum
3 months	0.95	0.07653785	0.0938558
	1	0.02793458	0.0625706
	1.05	0.01396729	0.0312853
6 months	0.95	0.1023344	0.1191295
	1	0.0550888	0.0822166
	1.05	0.0318363	0.0626412
1 year	0.95	0.1517857	0.1714171
	1	0.1071429	0.1383014
	1.05	0.0740133	0.1136681
2 years	0.95	0.2426658	0.2610702
	1	0.2028061	0.2318204
	1.05	0.1655976	0.2042068

Table 1: Bounds on the price of a European call option.

3.4.2. Asian call option

An arithmetic Asian call option with exercice date N, exercice price K and M averaging dates generates a pay-off $\left(\frac{1}{M}\sum_{i=1}^{M-1} S_{N-i} - K\right)_+$. This contingent claim is traded at time 0 for a price

$$\frac{1}{(1+r)^N}\mathbb{E}_{\widetilde{\mathbb{P}}}\left[\left(\frac{1}{M}\sum_{i=1}^{M-1} S_{N-i} - K\right)_+\right]$$

where $\widetilde{\mathbb{P}}$ is in the set $\mathcal{P}$ of all risk-neutral probability measures. For a comprehensive background on Asian option prices, the reader is referred, e.g., to Simon et al. (2000), Vanmaele et al. (2006) and the references therein.

As a numerical illustration, we consider the same parameter values as for the European call. Moreover, the averaging dates are taken to be all the dates during the life of the option (including maturity), i.e. $M = N$. Results are displayed in Table 2. The bounds displayed in Table 2 are computed by simulation using $10\,000$ random generations (standard errors attached to these approximations are also given). Again, the intervals of admissible prices is not too large.

N	K	Minimum	Std Error	Maximum	Std Error
6 months	0.95	0.08853629	0.0487%	0.1018841	0.0847%
	1	0.04004108	0.0483%	0.07083054	0.0660%
	1.05	0.01814324	0.0309%	0.04076772	0.0488%
1 year	0.95	0.1103681	0.0579%	0.1251327	0.1045%
	1	0.06580303	0.0585%	0.09112557	0.0926%
	1.05	0.0352885	0.0456%	0.06358178	0.0780%
2 years	0.95	0.152037	0.0733%	0.1629523	0.1391%
	1	0.1114088	0.0738%	0.1301361	0.1278%
	1.05	0.07566844	0.0684%	0.1022506	0.1161%

Table 2: Bounds on the price of an Asian call option.

4. DISCUSSION

In this paper, extremal elements in the class of risk-neutral probability measures are investigated, leading to bounds on the prices of contingent claims. This promising approach also leaves some open questions. It is well-known that improvements of the $\preceq_{\text{cx}}$-bounds are possible when the underlying distributions are unimodal (and are given by mixtures of uniform distributions). See, e.g., Denuit et al. (1999a). Unimodality is often satisfied under the physical probability measure. An interesting question could be to investigate the possible transmission of unimodality to the class of risk-neutral distributions. The same problem could be investigated with ageing notions.

Of course, alternative approaches could be investigated. For instance, convex bounds on the conditional distributions could be derived. From the definition of the process $\{X_n,\ n = 1, 2, \ldots\}$ we see that $\mathbb{E}[X_{n+1}|X_1, \ldots, X_n] = X_n\mathbb{E}[Y_{n+1}]$. If $\mathbb{E}[Y_n] = 1$ for all n (as it is usually the case in the financial applications, after a suitable change of measure) then $\mathbb{E}[X_{n+1}|X_1, \ldots, X_n] = X_n$. The idea is then to construct the extremal processes from the extremal conditional distributions from the knowledge of the support (a, b) and the conditional mean X_n.

ACKNOWLEDGEMENTS

Both authors gratefully acknowledge the financial support of the *Communauté française de Belgique* under contract "Projet d'Actions de Recherche Concertées" ARC 04/09-320. Michel Denuit thanks the *Banque Nationale de Belgique* for financial support under grant "Risk measures and Economic capital".

References

C. Courtois and M. Denuit. Bounds on convex reliability functions with known first moments. *European Journal of Operations Research*, 2005. in press.

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

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

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

M. Denuit, Cl. Lefèvre, and S. Utev. Stochastic orderings of convex/concave-type on an arbitrary grid. *Mathematics of Operations Research*, 24:835–846, 1999c.

J.C. Hull. *Options, Futures and Other Derivatives*. Prentice Hall, 5th edition, 2002.

G. Koshevoy and K. Mosler. The Lorenz zonoid of a multivariate distribution. *Journal of the American Statistical Association*, 91:873–882, 1996.

G. Koshevoy and K. Mosler. Multivariate Gini indices. *Journal of Multivariate Analysis*, 60: 252–276, 1997.

G. Koshevoy and K. Mosler. Lift zonoids, random convex hulls and the variability of random vectors. *Bernoulli*, 3:377–399, 1998.

S. Simon, M.J. Goovaerts, and J. Dhaene. An easy computable upper bound for the price of an arithmetic Asian option. *Insurance: Mathematics and Economics*, 26:175–183, 2000.

M. Vanmaele, G. Deelstra, J. Liinev, J. Dhaene, and M.J. Goovaerts. Bounds for the price of discretely sampled arithmetic Asian options. *Journal of Computational Mathematics*, 185(1): 51–90, 2006.

ON THE DIFFERENT APPROACHES FOR CAPITAL ALLOCATION

João Garcia[1], Thomas Alderweireld and Luc Léonard

Portfolio Management and Modelling Department, DEXIA S.A., Brussels, Belgium
Email: João.Garcia@dexia.com, Thomas.Alderweireld@dexia.com,
Luc.Léonard@dexia.com

Abstract

The concept of economic capital (EC) refers to the amount of capital a financial institution is supposed to set aside in order to prevent that its net asset value fall below a certain 'catastrophic level'. One then associates EC with the idea of a protection buffer for unexpected losses that might be incurred by the conglomerate. Traditionally one defines the EC with a certain confidence level (say 99.95%) of the loss distribution. The problem we will address is the computation of the total credit risk component of the economic capital and how to allocate it among the different entities of a financial conglomerate. The entities might be seen as business lines, portfolios or even whole institutions of a financial conglomerate. The model used to generate the loss distribution uses Monte Carlo (MC) simulation. For the time being, there are several models currently available in the literature for the allocation of the EC of a conglomerate among its different Business Lines. In this presentation we will present numerical results of a comparison between the main approaches, enlightening the drawbacks and advantages of each of them.

1. INTRODUCTION

The concept of *economic capital* (EC) refers to the amount of capital a financial institution is supposed to set aside in order to prevent its net asset value falling below a certain level that would have an impact on its normal operation. It is supposed to function as a buffer for any unexpected losses (ULs) that might be incurred by the institution.

On the regulatory side the Basel II framework has forced banks to use methodologies that link EC allocation techniques with risk. In addition, supervisors will be closely monitoring the procedures the banks will have put in place to deal with economic capital on Banking Supervision (BCBS). Moreover the increasing competition and its pressure on business margins have

[1] corresponding author

brought up the problem of efficient EC allocation among the different entities and business lines of a financial conglomerate. Directly related to the problem of efficient EC allocation is the problem of measuring the diversification benefits and risk adjusted contributions for individual positions and business lines taking into account the whole portfolio.

Several approaches have been proposed to calculate risk contributions at position and entity levels. Tasche (2004) for example shows that the only suitable way to measure performance is by defining a risk contribution as the derivative of the risk measure in the direction of the asset weight. Kalkbrener et al. (2004) on the other hand compares the expected shortfall measure with the classically used $VaR/CoVar$ approach. A derivation of EC allocation and risk measures is the paper of Dhaene et al. (2003). Additionally issues relating comonotonicity and EC allocation is treated in Dhaene et al. (2004). Recently Goovaerts et al. (2005) proposed an algorithm in which one uses both the whole loss distribution of the portfolio and the standalone distributions of the individual sub-portfolios to allocate EC. The problem this paper proposes to address is the one of comparing some of the different methodologies largely used in practice by market participants for allocating EC among the different entities of a financial conglomerate. Given the complexity of the task behind this work this paper should be seen as the first on a series of research articles in which the end goal is to present an approach for using such a system for active credit ALM portfolio management. In this scope a full discussion of the methodologies with a detailed explanation of the differences on the portfolios and positions levels is out of the scope of this article.

The paper will be structured as follows. In section 2 and section 3 we describe the portfolio model used and some of the different risk measures used in the market respectively. In section 4 we test those measures in a typical banking portfolio and give comments on the differences. In section 5 we conclude the article with a résumé of the differences and give a hint of the results of the forthcoming research.

2. THE PORTFOLIO MODEL

In order to be as realistic as possible with what is done in practice we have made our analysis on a one-period framework using a model (see Gupton et al. (1997) for details) commonly adopted by practitioners that also includes ratings migrations. In such a model the credit portfolio will consist of bonds whose returns Y_i are given by:

$$Y_i = \alpha \sum_{j=1}^{2} \beta_j Z_j + \sqrt{1-\alpha^2}\xi_i \tag{1}$$

with α representing the average correlation between the bonds and the systematic risk factors ((Z_1, Z_2): the market) which we suppose to be two: an industry and a country; and ξ_i being the idiosyncratic risk term ($\mathcal{N}(0,1)$: unidimensional gaussian distributed with mean zero and standard deviation one). The loss distribution of the portfolio will be given by:

$$L = \sum_{i=1}^{n} L_i \tag{2}$$

with the individual losses L_i given by one year forward changes in prices of each position due to rating migrations.

The rating migrations are determined using a standard Gaussian copula algorithm for the systematic factors and a standard uni-dimensional Gaussian for the idiosyncratic factor. The correlation between the market factors have been taken from the equity markets as explained in de Servigny and Renault (2003) [2]. The transition probability matrix (TPM) is a historical one and some standard adjustments have been done in order to compensate for some ratings incoherences.

3. ECONOMIC CAPITAL AND RISK MEASURES

Once the portfolio loss distribution has been determined one then uses risk measures for determining the unexpected loss and the allocation of the economic capital.

Assume a loss distribution defined by L and a certain quantile α. The credit value at risk ($cVaR$) and the expected shortfall (ES) associated with the quantile are defined respectively as:

$$cVaR_\alpha(L) = \inf\{x > 0 | P(L \leq x) \geq \alpha\} \tag{3}$$

$$ES_\alpha(L) = \mathbb{E}_P[L | L > cVaR_\alpha(L)]. \tag{4}$$

Below we will define approaches using standard risk measures that are largely used in practice for allocating economic capital. Assume the conglomerate is comprised of n sub-portfolios whose allocations we want to determine. The approaches we will be comparing in this paper are the following:

a) $VaR/CoVar$: although largely used by practitioners this approach is typical for the case of Gaussian loss distribution. In this approach the allocated capital of a certain sub-portfolio will be given by:

$$EC_i(\alpha) = \frac{cov\{L, L_i\}}{\sigma_T^2} \cdot EC_T(\alpha) \tag{5}$$

where L_i and L are the losses of sub and the total portfolio respectively. And σ_T^2 and $EC_T(\alpha)$ are the total portfolio loss variance and the EC for the total portfolio (assumed to be $cVaR_\alpha(L)$).

b) Pro-Rata $cVaR$: In this approach one uses the standalone $cVaR_\alpha$ of each sub-portfolio as a weight in the allocation of the total risk[3]:

$$EC_i(\alpha) = \frac{cVaR_\alpha(L_i)}{\sum cVaR_\alpha(L_i)} cVaR_\alpha(L). \tag{6}$$

c) Basel II: in this approach we use the relative proportions resulted from the Basel II formulas to allocate $cVaR$. Assume for example that Bsl_i is the regulatory capital of portfolio i then

[2] In our case we have used the equity correlations given by Portfolio Risk Tracker (PRT) from S&P.

[3] As measured by the total $cVaR$ that takes into account the whole correlation structure of the portfolio

the allocated capital for portfolio i will be given by:

$$EC_i(\alpha) = cVaR_T * Bsl_i/(\sum_{i=1}^{n} Bsl_i). \tag{7}$$

The principle behind this approach is to keep Basel II proportions for EC allocation.

d) Marginal Optimization of Total $cVaR_\alpha$ (see Goovaerts et al. (2005)): the idea is to search on the standalone loss distribution of each sub-portfolio the quantile for which the addition of the $cVaR_\alpha$ of the sub-portfolios would equal the total $cVaR$ of the whole portfolio. One then searches the quantile β on the standalone loss distribution of the sub-portfolios such that:

$$\beta = \inf\{\beta' \in [0,1] : \sum_{i=1}^{n} cVaR_{\beta'}(L_i) \geq cVaR_\alpha(L)\} \tag{8}$$

then $EC_i(\alpha)$ is defined as

$$EC_i(\alpha) = cVaR_\beta(L_i). \tag{9}$$

e) Credit VaR Contribution via Expected Shortfall: In this approach $cVaR$ is allocated using the concept of Expected Shortfall contribution. The Expected shortfall contribution is defined by:

$$ES_\beta(L_i) = E_P[L_i | L_i > cVaR_\beta(L_i)]. \tag{10}$$

The allocation is then given by:

$$EC_i(\alpha) = \frac{ES_\beta(L_i)}{ES_\beta(L)} cVaR_\alpha(L). \tag{11}$$

Observe that the quantiles for the $cVaR_\alpha$ and for the ES_β do not need to be the same. For example a bank might have its $cVaR_\alpha$ depending on a quantile α of (say) 99.97% while allocating it following a quantile β of 99%. I.e. portfolios that are more risky would need more capital. Such decisions are strategic and depend on the policy of the bank.

f) Expected Shortfall that equals $cVaR_\alpha$: in this approach we will be looking at the ES quantile that equals the $cVaR$. Then the allocation will be done using the ES. Assume for example that:

$$\beta = \inf\{\beta' \in [0,1] : ES_{\beta'}(L) \geq cVaR_\alpha(L)\}. \tag{12}$$

In this way:

$$EC_i(\alpha) = ES_\beta(L_i) \tag{13}$$

Observe that the main objective of this approach is to eliminate the problem that $cVaR$ is a non-additive measure (see Artzner et al. (1999) for details).

The results of the experiment will be given in function of the *diversification benefit (DB)* of a portfolio and it is defined as:

$$DB_T = 1 - EC_T/(\sum_{i=1}^{n} EC_i) \tag{14}$$

where EC_T is the total economic capital for the whole portfolio and $cVaR_\alpha(L_i)$ is the standalone $cVaR$ of the portfolio i composed by n sub-portfolios. The DB is a measure of the diversification gain one has when the sub-portfolios are put together in one large portfolio.

4. THE RESULTS

For the tests that follow we have selected 5 portfolios of different sizes, compositions and concentrations. The portfolios chosen have in general good quality and we have made them quite concentrated to show problems practitioners may face. The compositions of the different portfolios in terms of average rating, average maturity and concentration factor (defined as the percentage of issuers with 50% of the portfolio) are shown in table 1. We show the standalone $cVaR_{99.97\%}$ of each portfolio (it is given as a percentage of the total $cVaR_{99.97\%}$ of the whole portfolio).

In terms of sector concentrations we have build the portfolios P1 up to P5 with securities from mainly four sectors (financials, sovereigns, utilities, and some ABS's (not more than 10% of this class)) while portfolio P6 contains ABS's only. When comparing the sub-portfolios, P6 is the most diversified sub-portfolio.

	Avg Dur(yr)	Rating	CF	Amount(%)	Std-Alone VaR
P1	13	AA	5	43.3	36.5
P2	7	A+	7	35.4	56.3
P3	11	A+	11	5.4	13.3
P4	1	AA-	9	10.3	15.6
P5	30	AA	3	5.1	8.7
P6	5	AAA	30	0.5	0.4

Table 1: Composition of the Different Portfolios.

In the present analysis, one used the Moody's transition probability matrix adjusted for some rating imperfections. The correlation function is the one derived using equity returns. The correlation is the one that comes from PRT (Portfolio Risk Tracker)), a credit risk system comercialized by S&P (see de Servigny and Renault (2003) for more details). The systematic factor used, was calculated via regression using equity data and in this study we will be using 50% for that factor (although the market factor has proved to be quite lower than 45% we have been using 50% for a question of prudence). The result of the tests (in terms of DB) for the different methodologies is shown in table 2.

	P1	P2	P3	P4	P5	P6
a) VaR/$CoVaR$	1.8	19.0	48.9	82.8	-0.10	83.3
b) ProRata VaR	23.5	23.5	23.5	23.5	23.5	23.5
c) Basel II	27.0	7.1	36.7	72.7	6.0	64.1
d) Marg. Opt.	15.0	18.7	28.9	46.1	15.7	20.7
e) $cVaR$ contr	11.6	13.3	48.7	75.6	6.1	76.2
f) ES contr	20.2	10.5	48.7	62.0	14.1	80.0

Table 2: Diversification benefit of EC allocation using different measures.

A first observation is about the $Var/CoVar$ approach. As already reported elsewhere it can lead to a capital allocation that is higher than its standalone VaR and in some very special cases even higher than the whole amount of the portfolio. An example of it can be seen in the case of sub-portfolio P5.

The simple ProRata approach has the characteristic of dividing equally the DB among the sub-portfolios independent of the correlation within the sub-portfolios. As it is seen for sub-portfolio P6 this approach can have a negative impact on small sub-portfolios that would present ideal diversification characteristics with respect to the remaining portfolio. The Basel II approach has also the characteristic of simplicity (as the numbers are anyway available in the internal systems of most banks). The problem with this approach is that the correlation underlying Basel II formulas does not necessarily represent the correlation among the sub-portfolios under study (see e.g. the allocation given by the ES contrib. approach).

The Marginal Optimization uses the standalone loss distribution for allocation and that distribution does not take into account the correlation among the sub-portfolios. This is again evidenced by the allocation given to sub-portfolio P6. The $cVar$ and the ES contribution approaches account quite well the DB brought by sub-portfolio P6.

The $cVaR$ contribution methodology permits one to transfer risk among the sub-portfolios in a way that risk generated in low risk sub-portfolio is allocated to a higher risk one. Although it brings up the issue that the quantile used for allocation (99% in our case) is certainly arbitrary and certainly depends on management decisions. The ES contribution has the problem that one does not know in advance which quantile (β in the equation (12)) one will need to take, implying that one will need to make a couple of simulations to determine it (what can be time consuming).

The first three approaches have the advantage of simplicity at the cost of loosing important insights when allocating the DB. The Marginal Optimization method represents an increase in mathematical complexity. For communication purposes within subsidiaries and business lines it can be quite convenient: the standalone VaR is certainly available at the sub-portfolio levels and the holding would only need to pass the information on the specific quantile for allocation purpose. The disadvantage is the loss in correlation among the sub-portfolio when deciding the allocation. The $cVaR$ and the ES contributions both use ES factors for allocation purposes. The $cVar$ contribution brings up the quite (politically) sensitive issue of determining the quantile for allocation purpose. The ES contribution has the additional complexity of needing preliminary simulations to determine the allocation quantile.

5. CONCLUSIONS

In this paper we have shown the impact of different capital allocation methodologies for sub-portfolios of a large conglomerate and for individual positions. We have discussed six methodologies, three quite simple and straightforward: the $VaR/CoVar$, the ProRata and the Basell II (factors); and three more complex ones: Marginal Optimization, $cVaR$ contribution and ES contribution. The methodologies were tested on the problem of allocating risk as measured by $cVaR$ at the 99.97% quantile on six sub-portfolios.

The first three approaches have the advantage of simplicity at the cost of loosing important insights when allocating the DB. The Marginal Optimization method represents an increase in mathematical complexity. For communication purposes within subsidiaries and business lines it can be quite convenient: the standalone $cVaR$ is certainly available at the sub-portfolio level and the holding would only need to pass the information on the specific quantile for allocation purpose. The disadvantage is the loss in correlation among the sub-portfolio when deciding the allocation. The $cVaR$ and the ES contributions both use ES factors for allocation purposes. The $cVar$ contribution brings up the quite (politically) sensitive issue of determining the quantile for allocation purpose. The ES contribution has the additional complexity of needing preliminary simulations to determine the allocation quantile.

A continuation of this study includes building tables showing economic capital consumptions per position, rating, and sector. Those tables are put in the context of a portfolio management approach to an ALM credit desk. In order to avoid loosing resolution on a position level and being able to build a scenario analysis framework for the whole portfolio of the financial conglomerate a paralel system (with up to 25 machines) has been put in place being able to handle large amounts of positions in very short time. Additionally an innovative importance sampling algorithm has been implemented to improve the performance of the system. This study which represents the continuation of what has been shown in this paper will be published in brief.

References

P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.

A. de Servigny and O. Renault. Correlations evidence. *Risk*, 16(7):90–94, 2003.

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

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

M.J. Goovaerts, E. Van den Borre, and R.L.A. Laeven. Managing economic and virtual economic capital within financial conglomerates. *K.U.Leuven, Working Paper*, 2005.

G.M. Gupton, C.C. Finger, and M. Bhatia. Credit metrics – technical document. *Morgan Guaranty Trust*, 1997.

M. Kalkbrener, H. Lotter, and L. Overbeck. Sensible and efficient allocation for credit portfolios. *Risk*, 17(1):19–24, 2004.

Basel Committee on Banking Supervision (BCBS). International convergence of capital measurement and capital standards: a revised framework. 2004.

D. Tasche. Allocating portfolio economic capital to sub-portfolios. *Economic Capital: a Practioner Guide*, pages 275–302, 2004.

MINIMIZING THE (CONDITIONAL) VALUE-AT-RISK FOR A COUPON-BEARING BOND USING A BOND PUT OPTION.

Dries Heyman[†], Jan Annaert[†], Griselda Deelstra[‡] and Michèle Vanmaele[§]

[†]*Department of Financial Economics, Ghent University, Woodrow Wilsonplein 5/D, 9000 Gent, Belgium*
[‡]*Department of Mathematics, ISRO and ECARES, Université Libre de Bruxelles, CP 210, 1050 Brussels, Belgium*
[§]*Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281, S9, 9000 Gent, Belgium*
Email: dries.heyman@ugent.be, jan.annaert@ugent.be,
griselda.deelstra@ulb.ac.be, michele.vanmaele@ugent.be

Abstract

In this paper, we elaborate a formula for determining the optimal strike price for a bond put option, used to hedge a position in a bond. This strike price is optimal in the sense that it minimizes, for a given budget, either Value-at-Risk or Conditional Value-at-Risk. Formulas are derived for both zero-coupon and coupon bonds, which can also be understood as a portfolio of bonds. These formulas are valid for any short rate model with a given distribution of future bond prices.

1. INTRODUCTION

The importance of a sound risk management system can hardly be underestimated. The advent of new capital requirements for both the banking (Basel II) and insurance (Solvency II) industry, are two recent examples of the growing concern of regulators for the financial health of firms in the economy. This paper adds to this goal. In particular, we consider the problem of determining the optimal strike price for a bond put option, which is used to hedge the interest rate risk of an investment in a bond, zero-coupon or coupon-bearing. In order to measure risk, we focus on both Value-at-Risk and Conditional Value-at-Risk. Our optimization is constrained by a maximum hedging budget. Alternatively, our approach can also be used to determine the minimal budget a firm needs to spend in order to achieve a predetermined absolute risk level. This paper can be seen as an extension of Ahn et al. (1999), who consider the same problem for an investment in a share.

2. LOSS FUNCTION AND RISK MEASURES

Consider a portfolio with value W_t at time t. W_0 is then the value or price at which we buy the portfolio at time zero. W_T is the value of the portfolio at time T. The loss L we make by buying at time zero and selling at time T is then given by $L = W_0 - W_T$. The Value-at-Risk of this portfolio is defined as the $(1-\alpha)$-quantile of the loss distribution depending on a time interval with length T. A formal definition for the $\mathrm{VaR}_{\alpha,T}$ is

$$\Pr[L \geq \mathrm{VaR}_{\alpha,T}] = \alpha. \tag{1}$$

In other words $\mathrm{VaR}_{\alpha,T}$ is the loss of the worst case scenario on the investment at a $(1-\alpha)$ confidence level at time T. It is also possible to define the $\mathrm{VaR}_{\alpha,T}$ in a more general way

$$\mathrm{VaR}_{\alpha,T}(L) = \inf\left\{Y \mid \Pr(L > Y) \leq \alpha\right\}. \tag{2}$$

Although frequently used, VaR has attracted some criticisms. First of all, a drawback of the traditional Value-at-Risk measure is that it does not care about the tail behaviour of the losses. In other words, by focusing on the VaR at, let's say a 5% level, we ignore the potential severity of the losses below that 5% threshold. This means that we have no information on how bad things can become in a real stress situation. Therefore, the important question of 'how bad is bad' is left unanswered. Secondly, it is not a coherent risk measure, as suggested by Artzner et al. (1999). More specifically, it fails to fulfil the subadditivity requirement which states that a risk measure should always reflect the advantages of diversifying, that is, a portfolio will risk an amount no more than, and in some cases less than, the sum of the risks of the constituent positions. It is possible to provide examples that show that VaR is sometimes in contradiction with this subadditivity requirement.

Artzner et al. (1999) suggested the use of Conditional VaR (CVaR) as risk measure, which they describe as a coherent risk measure. CVaR is also known as TVaR, or Tail Value-at-Risk and is defined as follows:

$$\mathrm{CVaR}_{\alpha,T}(L) = \frac{1}{\alpha}\int_0^{\alpha} \mathrm{VaR}_{\beta,T}(L)\, d\beta. \tag{3}$$

This formula boils down to taking the arithmetic average of the quantiles of our loss, from 0 to α on, where we recall that $\mathrm{VaR}_{\beta,T}$ stands for the quantile at the level $1-\beta$, see (1). This formula already makes clear that $\mathrm{CVaR}_{\alpha,T}(L)$ will always be larger than $\mathrm{VaR}_{\alpha,T}(L)$.
If the cumulative distribution function of the loss is continuous, CVaR is also equal to the Conditional Tail Expectation (CTE) which for the loss L is calculated as:

$$\mathrm{CTE}_{\alpha,T}(L) = \mathrm{E}[L \mid L > \mathrm{VaR}_{\alpha,T}(L)].$$

3. THE BOND HEDGING PROBLEM

Analogously to Ahn et al. (1999), we assume that we have, at time zero, one bond with maturity S and we will sell this bond at time T, which is prior to S. In case of an increase in interest rates, not hedging can lead to severe losses. Therefore, the company decides to spend an amount C on hedging. This amount will be used to buy one or part of a bond put option, so that, in case of a substantial decrease in the bond price, the put option can be exercised in order to prevent large losses. The remaining question now is how to choose the strike price. We will find the optimal strike prices which minimize VaR and CVaR respectively for a given hedging cost. An alternative interpretation of our setup is that it can be used to calculate the minimal hedging budget the firm has to spend in order to achieve a specified VaR or CVaR level. The latter setup was followed in the paper by Miyazaki (2001).

3.1. Zero-coupon bond

Let us assume that the institution has an exposure to a bond, $Y(0,S)$, with principal $K = 1$, which matures at time S, and that the company has decided to hedge the bond value by using a percentage h ($0 < h < 1$) of one put option $P(0,T,S,X)$ with strike price X and exercise date T (with $T \leq S$).
Further, we assume that the distribution of $Y(T,S)$ is known and is continuous and strictly increasing. We will denote its cumulative distribution function (cdf) under the measure in which we measure the VaR or the CVaR by $F_{Y(T,S)}(\cdot)$. For example when the short-rate model is one of the following commonly used interest rate models such as Vasicek, one- and two-factor Hull-White, two-factor additive Gaussian model G2++, two-factor Heath-Jarrow-Morton with deterministic volatilities, see e.g. Brigo and Mercurio (2001), then $Y(T,S)$ has a lognormal distribution.

Analogously to Ahn et al. (1999), we can look at the future value of the hedged portfolio that is composed of the bond Y and the put option $P(0,T,S,X)$ at time T as a function of the form

$$H_T = \max(hX + (1-h)Y(T,S), Y(T,S)).$$

In a worst case scenario — a case which is of interest to us — the put option finishes in-the-money. Then the future value of the portfolio equals

$$H_T = (1-h)Y(T,S) + hX.$$

Taking into account the cost of setting up our hedged portfolio, which is given by the sum of the bond price $Y(0,S)$ and the cost C of the position in the put option, we get for the value of the loss:

$$L = Y(0,S) + C - ((1-h)Y(T,S) + hX), \tag{4}$$

and this under the assumption that the put option finishes in-the-money.
Note that this loss function can be seen as a strictly decreasing function f in $Y(T,S)$:

$$f(Y(T,S)) := Y(0,S) + C - ((1-h)Y(T,S) + hX). \tag{5}$$

VaR minimization
We first look at the case of determining the optimal strike X when minimizing the VaR under a constraint on the hedging cost.

Recalling (1) and (4), the Value-at-Risk at an α percent level of a position $H = \{Y, h, P\}$ consisting of a bond Y and h put options P (which are assumed to be in-the-money at expiration) with a strike price X and an expiry date T is equal to[1]

$$\mathrm{VaR}_{\alpha,T}(L) = Y(0,S) + C - ((1-h)F^{-1}_{Y(T,S)}(\alpha) + hX), \tag{6}$$

where $F^{-1}_{Y(T,S)}(\alpha)$ is the percentile of the cdf $F_{Y(T,S)}$, i.e. $\Pr[Y(T,S) \leq F^{-1}_{Y(T,S)}(\alpha)] = \alpha$.

Similar to the Ahn et al. problem, we would like to minimize the risk of the future value of the hedged bond H_T, given a maximum hedging expenditure C. More precisely, we consider the minimization problem

$$\min_{X,h} Y(0,S) + C - ((1-h)F^{-1}_{Y(T,S)}(\alpha) + hX)$$

subject to the restrictions $C = hP(0,T,S,X)$ and $h \in (0,1)$.
This is a constrained optimization problem with Lagrange function

$$\mathcal{L}(X,h,\lambda) = \mathrm{VaR}_{\alpha,T}(L) - \lambda(C - hP(0,T,S,X)),$$

containing one multiplicator λ. Note that the multiplicators to include the inequalities $0 < h$ and $h < 1$ are zero since these constraints are not binding. Taking into account that the optimal strike X^* will differ from zero, we find from the Kuhn-Tucker conditions

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial X} = -h + h\lambda \dfrac{\partial P}{\partial X}(0,T,S,X) = 0 \\ \dfrac{\partial \mathcal{L}}{\partial h} = -(X - F^{-1}_{Y(T,S)}(\alpha)) + \lambda P(0,T,S,X) = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = C - hP(0,T,S,X) = 0 \\ 0 < h < 1 \qquad \text{and} \qquad \lambda > 0 \end{cases}$$

that this optimal strike X^* should satisfy the following equation

$$P(0,T,S,X) - (X - F^{-1}_{Y(T,S)}(\alpha))\frac{\partial P}{\partial X}(0,T,S,X) = 0. \tag{7}$$

By a change of numeraire, it is well known that the put option price equals the discounted expectation under the T-forward measure of the the pay-off:

$$P(0,T,S,X) = Y(0,T)\mathrm{E}^T[(X - Y(T,S))_+].$$

Its first order derivative with respect to the strike X gives the cumulative distribution function $F^T_{Y(T,S)}$ of $Y(T,S)$ under this T-forward measure, see Breeden and Litzenberger (1978):

$$\frac{\partial P}{\partial X}(0,T,S,X) = Y(0,T)F^T_{Y(T,S)}(X). \tag{8}$$

Hence, (7) is equivalent to

$$P(0,T,S,X) - (X - F^{-1}_{Y(T,S)}(\alpha))Y(0,T)F^T_{Y(T,S)}(X) = 0.$$

[1] In case of an unhedged portfolio, take $C = h = 0$ in (4) and in (6) to obtain the loss function L with corresponding $\mathrm{VaR}_{\alpha,T}(L)$.

Important remarks

1. We note that the optimal strike price is independent of the hedging cost C. This independence implies that for the optimal strike X^*, VaR in (6) is a linear function of h (or C):

$$\mathrm{VaR}_{\alpha,T}(L) = Y(0,S) - F^{-1}_{Y(T,S)}(\alpha) + h(P(0,T,S,X^*) + F^{-1}_{Y(T,S)}(\alpha) - X^*).$$

   So, there is a linear trade-off between the hedging expenditure and the VaR level. It is a decreasing function since in view of (8) $\frac{\partial P}{\partial X}(0,T,S,X^*) < 1$ and thus according to (7) $X^* - F^{-1}_{Y(T,S)}(\alpha) > P(0,T,S,X^*)$.
   Although the setup of the paper is determining the strike price which minimizes a certain risk criterion, given a predetermined hedging budget, this trade-off shows that the analysis and the resulting optimal strike price can evidently also be used in the case where a firm is fixing a nominal value for the risk criterion and seeks the minimal hedging expenditure needed to achieve this risk level. It is clear that, once the optimal strike price is known, we can determine, in both approaches, the remaining unknown variable (either VaR, either C).

2. We also note that the optimal strike price is higher than the bond VaR level $F^{-1}_{Y(T,S)}(\alpha)$. This has to be the case since $P(0,T,S,X)$ is always positive and the change in the price of a put option due to an increase in the strike is also positive. This result is also quite intuitive since there is no point in taking a strike price which is situated below the bond price you expect in a worst case scenario.
   When moreover the optimal strike is smaller than the forward price of the bond, i.e.

$$X^* < \frac{Y(0,S)}{Y(0,T)},$$

   then the price of put option to buy will be small.

3. The assumption of continuity and strictly monotonicity of the distribution of $Y(T,S)$ can be weakened. In that case we should work with the general definition (2) of VaR.

CVaR minimization

In this section, we demonstrate the ease of extending our analysis to the alternative risk measure CVaR (3) by integration of (6):

$$\mathrm{CVaR}_{\alpha,T}(L) = Y(0,S) + C - hX - \frac{1}{\alpha}(1-h)\int_0^{\alpha} F^{-1}_{Y(T,S)}(\beta)d\beta. \tag{9}$$

We again seek to minimize this risk measure, in order to minimize potential losses. The procedure for minimizing this CVaR is analogue to the VaR minimization procedure. The resulting optimal strike price X^* can thus be determined from the implicit equation below:

$$P(0,T,S,X) - (X - \frac{1}{\alpha}\int_0^{\alpha} F^{-1}_{Y(T,S)}(\beta)d\beta)\frac{\partial P}{\partial X}(0,T,S,X) = 0, \tag{10}$$

or, equivalently by (8), from

$$P(0,T,S,X) - (X - \frac{1}{\alpha}\int_0^{\alpha} F^{-1}_{Y(T,S)}(\beta)d\beta)Y(0,T)F^{T}_{Y(T,S)}(X) = 0.$$

As for the VaR-case the optimal strike X^* is independent of the hedging cost C and CVaR can be plotted as a linear function of C (or h) representing a trade-off between the cost and the level of protection.
For the same reason as in the VaR-case, the optimal strike X^* has to be higher than the bond CVaR level $\frac{1}{\alpha}\int_0^\alpha F^{-1}_{Y(T,S)}(\beta)d\beta$.

4. COUPON-BEARING BOND

We consider now the case of a coupon-bearing bond paying cash flows $\mathcal{C} = [c_1, \ldots, c_n]$ at maturities $\mathcal{S} = [S_1, \ldots, S_n]$. Let $T \leq S_1$. The price of this coupon-bearing bond in T is expressed as a linear combination (or a portfolio) of zero-coupon bonds:

$$\text{CB}(T, \mathcal{S}, \mathcal{C}) = \sum_{i=1}^{n} c_i Y(T, S_i). \tag{11}$$

As in the previous section, the company wants to hedge its position in this bond by buying a percentage of a put option on this bond with strike X and maturity T. In order to determine the strike X, the VaR or the CVaR of the hedged portfolio at time T is minimized under a budget constraint. Comparing the results in the previous section for VaR and CVaR minimization for a hedged position in zero-coupon bond we note that both cases can in fact be treated together.

We first have a look at the value of a put option on a coupon-bearing bond as well as at the structure of the loss function.
Since the zero-coupon bonds $Y(T, S_i)$ all depend on the same short rate at T, the vector $(Y(T, S_1),$ $\ldots,Y(T, S_n))$ is comonotonic, see Kaas et al. (2000). By the properties of comonotonic vectors, the coupon-bearing bond $\text{CB}(T, \mathcal{S}, \mathcal{C})$ (11) is a comonotonic sum with cumulative distribution function $F^T_{\text{CB}}(\cdot)$ under the T-forward measure. This implies that a European option on a coupon-bearing bond decomposes into a portfolio of options on the individual zero-coupon bonds in the portfolio, which gives in case of a put with maturity T and strike X:

$$\text{CB}P(0, T, \mathcal{S}, \mathcal{C}, X) = \sum_{i=1}^{n} c_i P(0, T, S_i, X_i), \quad \text{with} \quad \sum_{i=1}^{n} c_i X_i = X. \tag{12}$$

This result, now well-known as the Jamshidian decomposition, was found in Jamshidian (1989) in case of a Vasicek interest rate model. Kaas et al. (2000) obtained this result in a more general framework of stop-loss premiums and gave an explicit expression for the X_i:

$$X_i = (F^T_{Y(T,S_i)})^{-1}(F^T_{\text{CB}}(X)). \tag{13}$$

Repeating the reasoning of Section 3.1 we may conclude that in a worst case scenario the loss of the hedged portfolio at time T composed of the coupon-bearing bond (11) and the put option (12) equals a strictly decreasing function f of the random variable $\text{CB}(T, \mathcal{S}, \mathcal{C})$:

$$L = \text{CB}(0, \mathcal{S}, \mathcal{C}) + C - ((1-h)\text{CB}(T, \mathcal{S}, \mathcal{C}) + hX) := f(\text{CB}(T, \mathcal{S}, \mathcal{C})). \tag{14}$$

VaR and CVaR minimization

The VaR of this loss that we want to minimize under the constraints $0 < h < 1$ and $C = h\text{CB}P(0, T, \mathcal{S}, \mathcal{C}, X)$, is given by

$$\text{VaR}_{\alpha,T}(L) = f(F_{\text{CB}}^{-1}(\alpha)) = \text{CB}(0, \mathcal{S}, \mathcal{C}) + C - ((1-h)F_{\text{CB}}^{-1}(\alpha) + hX), \tag{15}$$

where F_{CB}^{-1} stands for the inverse cdf of the coupon-bearing bond under the measure in which VaR (and CVaR) is measured.

By integrating this relation (15), after replacing α by β, with respect to β between the integration bounds 0 and α, we find for the CVaR of the loss:

$$\text{CVaR}_{\alpha,T}(L) = \text{CB}(0, \mathcal{S}, \mathcal{C}) + C - hX - \frac{1}{\alpha}(1-h)\int_0^{\alpha} F_{\text{CB}}^{-1}(\beta)d\beta. \tag{16}$$

Also here we note the similarity in the expressions for the risk measures (RM) VaR and CVaR which could be collected in one expression:

$$\text{RM}_{\alpha,T}(L) = \text{CB}(0, \mathcal{S}, \mathcal{C}) + C - hX - (1-h)g(F_{\text{CB}}^{-1}(\alpha)) \tag{17}$$

$$\text{with} \quad g(F_{\text{CB}}^{-1}(\alpha)) = \begin{cases} F_{\text{CB}}^{-1}(\alpha) & \text{if RM} = \text{VaR} \\ \dfrac{1}{\alpha}\displaystyle\int_0^{\alpha} F_{\text{CB}}^{-1}(\beta)d\beta & \text{if RM} = \text{CVaR}. \end{cases} \tag{18}$$

Although the marginal distributions $F_{Y(T,S_i)}$ are known, the distribution F_{CB} of the sum can in general not be obtained. However, in the case of a comonotonic sum we have, see again Kaas et al. (2000),

$$F_{\text{CB}}^{-1}(p) = \sum_{i=1}^{n} c_i F_{Y(T,S_i)}^{-1}(p) \qquad \text{for all } p \in [0,1], \tag{19}$$

and similarly for the inverse cdfs under the T-forward measure.

We now want to solve the constrained optimization problem

$$\min_{X,h} \text{RM}_{\alpha,T}(L) \quad \text{subjected to} \quad C = h\text{CB}P(0, T, \mathcal{S}, \mathcal{C}, X), \quad 0 < h < 1.$$

From the Kuhn-Tucker conditions we find that the optimal strike price X^* satisfies the following equation

$$\text{CB}P(0, T, \mathcal{S}, \mathcal{C}, X) - (X - g(F_{\text{CB}}^{-1}(\alpha)))\frac{\partial \text{CB}P}{\partial X}(0, T, \mathcal{S}, \mathcal{C}, X) = 0. \tag{20}$$

Rewriting this equation in terms of the put options on the individual zero-coupon bonds cfr. (12), invoking (19) and using the linearity of the function g (18), leads to the following equivalent set of equations:

$$\sum_{i=1}^{n} c_i P(0, T, S_i, X_i) - (X - \sum_{i=1}^{n} c_i g(F_{Y(T,S_i)}^{-1}(\alpha)))\sum_{i=1}^{n} c_i \frac{\partial P}{\partial X_i}(0, T, S_i, X_i)\frac{\partial X_i}{\partial X} = 0 \tag{21}$$

$$\sum_{i=1}^{n} c_i X_i = X \tag{22}$$

$$\sum_{i=1}^{n} c_i \frac{\partial X_i}{\partial X} = 1, \tag{23}$$

where X_i is defined by (13).
We can further simplify relation (21) by applying relation (8) to the strike X_i given by (13), i.e.

$$\frac{\partial P}{\partial X_i}(0,T,S_i,X_i) = Y(0,T)F^T_{Y(T,S_i)}((F^T_{Y(T,S_i)})^{-1}(F^T_{\text{CB}}(X))) = Y(0,T)F^T_{\text{CB}}(X).$$

Hence, this derivative is independent of i which implies in view of (23) that

$$\sum_{i=1}^{n} c_i \frac{\partial P}{\partial X_i}(0,T,S_i,X_i)\frac{\partial X_i}{\partial X} = Y(0,T)F^T_{\text{CB}}(X)\sum_{i=1}^{n} c_i \frac{\partial X_i}{\partial X} = Y(0,T)F^T_{\text{CB}}(X). \tag{24}$$

We introduce the short hand notation

$$A_X := F^T_{\text{CB}}(X). \tag{25}$$

Substitution of (13), (22) and (24) in (21) leads to the following equation that we have to solve for A_X:

$$\sum_{i=1}^{n} c_i P(0,T,S_i,(F^T_{Y(T,S_i)})^{-1}(A_X)) - Y(0,T)A_X \sum_{i=1}^{n} c_i[(F^T_{Y(T,S_i)})^{-1}(A_X) - g(F^{-1}_{Y(T,S_i)}(\alpha))] = 0. \tag{26}$$

Once, we know A_X we immediately have the optimal strike X^* from (22):

$$X^* = \sum_{i=1}^{n} c_i (F^T_{Y(T,S_i)})^{-1}(A_X). \tag{27}$$

Remarks

1. We note that also in the case of a coupon-bearing bond the optimal strike price is independent of the hedging cost and that one can look at the trade-off between the hedging expenditure and the RM level, cfr. Section 3.1.

2. Also here we may weaken the assumption of continuity and strictly monotonicity of the distribution functions $F_{Y(T,S_i)}$. In that case we have to invoke Kaas et al. (2000) with a so-called η-inverse distribution of a random variable Y which is defined as the following convex combination:

$$F_Y^{-1(\eta)}(p) = \eta F_Y^{-1}(p) + (1-\eta)F_Y^{-1+}(p), \quad p \in (0,1)\,,\ \eta \in [0,1],$$
$$F_Y^{-1}(p) = \inf\{y \in \mathbb{R} \mid F_Y(y) \geq p\}\,, \quad p \in [0,1]\,,$$
$$F_Y^{-1+}(p) = \sup\{y \in \mathbb{R} \mid F_Y(y) \leq p\}\,, \quad p \in [0,1]\,.$$

Thus relation (12) holds with

$$X_i = (F^T_{Y(T,S_i)})^{-1(\eta)}(F^T_{\text{CB}}(X)),$$

where $\eta \in [0,1]$ is determined from

$$\sum_{i=1}^{n} c_i (F^T_{Y(T,S_i)})^{-1(\eta)}(F^T_{\text{CB}}(X)) = X.$$

5. APPLICATION: HULL-WHITE MODEL

As an application, we focus on the Hull-White one-factor model, first discussed by Hull and White in 1990 (see Hull and White (1990)). We choose this model because it is still an often used model in financial institutions for risk management purposes, (see Brigo and Mercurio (2001)).

Hull and White (1990) assume under the risk-neutral measure Q that the instantaneous interest rate follows a mean reverting process also known as an Ornstein-Uhlenbeck process:

$$dr(t) = (\theta(t) - \gamma(t)r(t))dt + \sigma(t)dZ(t) \tag{28}$$

with $Z(t)$ a standard Brownian motion under Q, and with time dependent parameters $\theta(t)$, $\gamma(t)$, and $\sigma(t)$. The parameter $\theta(t)$ is the time dependent long-term average level of the spot interest rate around which $r(t)$ moves, $\gamma(t)$ controls the mean-reversion speed and $\sigma(t)$ is the volatility function. By making the mean reversion level θ time dependent, a perfect fit with a given term structure can be achieved, and in this way arbitrage can be avoided. In our analysis, we will keep γ and σ constant, and thus time-independent. According to Brigo and Mercurio (2001), this is desirable when an exact calibration to an initial term structure is wanted. This perfect fit then occurs when $\theta(t)$ satisfies the following condition:

$$\theta(t) = F_t^M(0,t) + \gamma F^M(0,t) + \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t}),$$

where, $F^M(0,t)$ denotes the instantaneous forward rate observed in the market on time zero with maturity t.

It can be shown (see Hull and White (1990)) that the expectation and variance of the stochastic variable $r(t)$ are:

$$\text{E}\left[r(t)\right] = m(t) = r(0)e^{-\gamma t} + a(t) - a(0)e^{-\gamma t}, \qquad \text{Var}\left[r(t)\right] = s^2(t) = \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t}) \tag{29}$$

with the expression $a(t)$ calculated as follows:

$$a(t) = F^M(0,t) + \frac{\sigma^2}{2}\left(\frac{1 - e^{-\gamma t}}{\gamma}\right)^2.$$

Based on these results, Hull and White developed an analytical expression for the price of a zero-coupon bond with maturity date S

$$Y(t,S) = A(t,S)e^{-B(t,S)r(t)},$$

where

$$B(t,S) = \frac{1 - e^{-\gamma(S-t)}}{\gamma}, \qquad A(t,S) = \frac{Y^M(0,S)}{Y^M(0,t)}e^{B(t,S)F^M(0,t) - \frac{\sigma^2}{4\gamma}(1-e^{-2\gamma t})B^2(t,S)}$$

with Y^M the bond price observed in the market. Since $A(t,S)$ and $B(t,S)$ are independent of $r(t)$, the distribution of a bond price at any given time must be lognormal with parameters Π and Σ^2:

$$\Pi(t,S) = \ln A(t,S) - B(t,S)m(t), \qquad \Sigma(t,S)^2 = B(t,S)^2 s^2(t),$$

with $m(t)$ and $s^2(t)$ given by (29). Thus under the risk neutral measure the inverse cdf of $Y(T,S)$ is given by

$$F^{-1}_{Y(T,S)}(p) = e^{\Pi(T,S)+\Sigma(T,S)\Phi^{-1}(p)}, \qquad p \in [0,1], \tag{30}$$

and we can compute the (standard) integral

$$\int_0^{\alpha} F^{-1}_{Y(T,S)}(\beta)d\beta = e^{\Pi(T,S)} \int_0^{\alpha} e^{\Sigma(T,S)\Phi^{-1}(\beta)}d\beta = e^{\Pi(T,S)+\frac{1}{2}\Sigma^2(T,S)}\Phi(\Phi^{-1}(\alpha) - \Sigma(T,S)). \tag{31}$$

By a change of numeraire it can be shown that $Y(T,S)$ remains lognormally distributed under the T-forward measure but now with parameters Π^T and $(\Sigma^T)^2$ given by:

$$\Pi^T(T,S) = \ln\left(\frac{Y(0,S)}{Y(0,T)}\right) - \frac{1}{2}(\Sigma^T(T,S))^2, \quad \Sigma^T(T,S) = \Sigma(T,S). \tag{32}$$

Hence, the inverse cdf of $Y(T,S)$ under the T-forward measure is known explicitly:

$$(F^T_{Y(T,S)})^{-1}(p) = e^{\Pi^T(T,S)+\Sigma(T,S)\Phi^{-1}(p)}, \qquad p \in [0,1], \tag{33}$$

as well as the put option price and its derivative with respect to the strike:

$$P(0,T,S,X) = -Y(0,S)\Phi(-d_1(X)) + XY(0,T)\Phi(-d_2(X)),$$
$$\frac{\partial P}{\partial X}(0,T,S,X) = Y(0,T)\Phi(-d_2(X)),$$

with, when taking (32) into account,

$$d_1(X) = \frac{1}{\Sigma(T,S)}\left[\ln\left(\frac{Y(0,S)}{Y(0,T)}\right) - \ln(X)\right] + \frac{1}{2}\Sigma(T,S) = \frac{\Pi^T(T,S) - \ln(X)}{\Sigma(T,S)} + \Sigma(T,S) \tag{34}$$

$$d_2(X) = d_1(X) - \Sigma(T,S) = \frac{\Pi^T(T,S) - \ln(X)}{\Sigma(T,S)}. \tag{35}$$

For the **zero-coupon case**, substitution of the relations above in (7) and in (10) gives the following implicit relation for the optimal strike X^*:

$$G(\Phi^{-1}(\alpha)) = \frac{Y(0,S)\Phi(-d_1(X))}{Y(0,T)\Phi(-d_2(X))},$$

with

$$G(\Phi^{-1}(\alpha)) = \begin{cases} e^{\Pi(T,S)+\Sigma(T,S)\Phi^{-1}(\alpha)} & \text{if VaR} \\ \frac{1}{\alpha}e^{\Pi(T,S)+\frac{1}{2}\Sigma^2(T,S)}\Phi(\Phi^{-1}(\alpha) - \Sigma(T,S)) & \text{if CVaR.} \end{cases} \tag{36}$$

For the **coupon-bearing bond case**, the above relations for the distribution and the put option price hold but with S and X replaced by S_i and X_i. The expressions (34) and (35) for $d_1(X_i)$ and $d_2(X_i)$ can further be simplified in view of (13),(25) and (31):

$$d_1(X_i) = \Sigma(T,S_i) - \Phi^{-1}(A_X), \qquad d_2(X_i) = -\Phi^{-1}(A_X).$$

In this way, the set of equations (26)-(27) to find the optimal strike X^* is equivalent with:

$$\sum_{i=1}^{n} c_i \left[-Y(0,S_i)\Phi(\Phi^{-1}(A_X) - \Sigma(T,S_i)) + Y(0,T)A_X e^{\Pi^T(T,S_i)+\Sigma(T,S_i)\Phi^{-1}(A_X)} \right]$$
$$= Y(0,T)A_X \sum_{i=1}^{n} c_i \left[e^{\Pi^T(T,S_i)+\Sigma(T,S_i)\Phi^{-1}(A_X)} - G_i(\Phi^{-1}(\alpha)) \right]$$
$$X^* = \sum_{i=1}^{n} c_i e^{\Pi^T(T,S_i)+\Sigma(T,S_i)\Phi^{-1}(A_X)},$$

where $G_i(\Phi^{-1}(\alpha))$ is defined by (36) when replacing S by S_i.

For a complete numerical example we refer to Deelstra et al. (2005) and Heyman et al. (2006).

6. CONCLUSIONS

We provided a method for minimizing the risk of a position in a bond (zero-coupon or coupon-bearing) by buying (a percentage of) a bond put option. Taking into account a budget constraint, we determine the optimal strike price, which minimizes a Value-at-Risk or Conditional Value-at-Risk criterion. Alternatively, our approach can be used when a nominal risk level is fixed, and the minimal hedging budget to fulfil this criterion is desired. From the class of short rate models which result in lognormally distributed future bond prices, we have selected the Hull-White one-factor model for an illustration of our optimization.

References

D.H. Ahn, J. Boudoukh, M. Richardson, and R. Whitelaw. Optimal risk management using options. *Journal of Finance*, 54:359–375, 1999.

P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–229, 1999.

D.T. Breeden and R.H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51:621–651, 1978.

D. Brigo and F. Mercurio. *Interest Rate Models — Theory and Practice.* Springer Verlag, Berlin, 2001.

G. Deelstra, A. Ezzine, D. Heyman, and M. Vanmaele. Managing Value-at-Risk for a bond using bond put options. *Proceedings ASMDA 2005, Brest,* pages 932–941, 2005.

D. Heyman, J. Annaert, G. Deelstra, and M. Vanmaele. Risk management of a bond portfolio using options. Technical report, Ghent University, 2006.

J. Hull and A. White. Pricing interest rate derivative securities. *Review of Financial Studies*, 3(4): 573–592, 1990.

F. Jamshidian. An exact bond option formula. *The Journal of Finance*, 44(1):205–209, 1989.

R. Kaas, J. Dhaene, and M. Goovaerts. Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, 27(2):151–168, 2000.

K. Miyazaki. Dual analysis on hedging VaR of bond portfolio using options. Technical report, University of Electro-Communications, Tokyo, 2001. URL `http://www.gloriamundi.org/picsresources/km.pdf`.

SEVERAL TWO-BOUNDARY PROBLEMS FOR LÉVY PROCESSES

Tetyana Kadankova and Noël Veraverbeke

Center for Statistics, Hasselt University, Agoralaan, Building D, 3590 Diepenbeek, Belgium
Email: `tetyana.kadankova@uhasselt.be, noel.veraverbek@uhasselt.be`

Abstract

Several two-sided exit problems for a Lévy process are considered in the present paper. We obtain the integral transforms of the joint distribution of the first exit time from a fixed interval and the value of the overshoot through boundaries at the instant of the first exit. The Laplace transform is found of the joint distribution of the number of upward and downward intersections. Finally, the joint distribution of the first entry time into a given interval and the value of the process at that instant is obtained.

1. INTRODUCTION

During the last decades Lévy processes have become very popular as modeling tools in insurance and mathematical finance (see i.e. Boyarchenko and Levendorskii (2002), Schoutens (2004), Asmussen et al. (2004)). Some specific types of Lévy processes proved to be appropriate as models of stock prices (Geman (2002), Rydberg (1997), Schoutens (2001), Asmussen et al. (2005)). It has been recognized that Lévy models give a much better fit to the financial data and lead to significant improvement with respect to the Black & Scholes model, see Schoutens (2003). Along with the applications aspect, the theory of Lévy processes itself has faced with a lot of developments (Asmussen and Rosinski (2001), Bertoin (1997), Pistorius (2004), Kadankov and Kadankova (2005), Kyprianou and Pistorius (2003), Avram et al. (2002) and many others). Many interesting problems in applied probability and in finance, in particular, are related to determining of the distribution of the first exit time and the value of the process at the epoch of exit. However, other boundary characteristics of the process are also of interest. Motivated by this fact, in this framework we consider several other boundary problems. The first problem we deal with is the so-called two-sided exit problem, which plays a crucial role in options pricing. We obtain the integral transforms of the joint distribution of the exit from the interval and the value of the overshoot through the boundaries. Further, employing these results we derive the exact formulae for the integral transforms of the joint distribution of the supremum, the infimum and the value of the Lévy process. The distribution of the number of intersections of the interval by a general Lévy process is obtained in terms of the integral transforms of the joint distribution of the exit from the interval and the value of the overshoot through the boundary, the first passage time and the value of the overshoot through the level. Finally, the joint distribution of the first entry time into the interval and the value of the process at this instant are determined in terms of integral transforms. Note however, that the exact formulae for the integral transforms of the mentioned functionals are obtained but not the distributions themselves, which we are primarily interested in. Thus, we are faced with a problem of inverting the integral transforms which is of high dimension. An alternative is the use of Monte Carlo simulations, which is not a simple task. Further simplifications of the obtained formulae is the subject of ongoing research.

2. DEFINITIONS AND AUXILIARY RESULTS

Let $(\Omega,\ \mathfrak{F}, \{\mathfrak{F}_t\}, P)$ be a filtered probability space, where the filtration $\{\mathfrak{F}_t\}$ satisfies the usual conditions of right-continuity and completion. We assume that all random variables and stochastic processes are defined on this probability space. A Lévy process is a $\mathfrak{F}$-adapted stochastic process $\{\xi(t);\ t \geq 0\}$ which has independent and stationary increments whose paths are right-continuous with left limits (see i.e. Sato (1999) or Bertoin (1996)). Under the assumption that $\xi(0) = 0$, the Laplace transform of the process $\{\xi(t);\ t \geq 0\}$ has the form $E[e^{-p\xi(t)}] = e^{t\,k(p)}$, $\operatorname{Re} p = 0$, where the function $k(p)$ is called the Laplace exponent and is given by the formula (Skorokhod (1971), p.110)

$$k(p) = \frac{1}{t}\ln E[e^{-p\xi(t)}] = \frac{1}{2}\,p^2\sigma^2 - \alpha p + \int_{-\infty}^{\infty} \left(e^{-px} - 1 + \frac{px}{1+x^2} \right) \Pi(dx). \tag{1}$$

Here $\alpha,\ \sigma \in \mathbb{R}$ and $\Pi(\cdot)$ is a measure on the real line, such that $\int_{-1}^{1} x^2\Pi(dx) < \infty$. The introduced process is a space homogeneous, strong Markov process. Let us fix $B > 0$ and define the variable

$$\chi(y) = \inf\{\, t:\ y + \xi(t) \notin [0, B]\,\}, \quad y \in [0, B]$$

the first exit time from the interval $[0, B]$ by the process $y + \xi(t)$. The random variable $\chi(y)$ is a Markov time and $P\,[\,\chi(y) < \infty\,] = 1$. Exit from the interval $[0, B]$ can take place either through the upper boundary B, or through the lower boundary 0. Introduce events: $A^B = \{w:\ \xi(\chi(y)) > B\,\}$, i.e. the exit takes place through the upper boundary; $A_0 = \{w:\ \xi(\chi(y)) < 0\,\}$, i.e. the exit takes place through the lower boundary. Define

$$X(y) = (\xi(\chi(y)) - B)\, I_{A^B} + (-\xi(\chi(y)))\, I_{A_0}, \qquad P\,[\,A^B + A_0\,] = 1$$

the value of the overshoot through one of the boundaries at the epoch of the exit, where $I_A = I_A(\omega)$ is the indicator of the event A. To determine the joint distribution of $\{\chi,\ X\}$ we will employ one-boundary characteristics of the process. For $x \geq 0$ introduce the random variables

$$\tau^x = \inf\{t : \xi(t) > x\}, \quad T^x = \xi(\tau^x) - x, \quad \tau_x = \inf\{t : \xi(t) < -x\,\}, \quad T_x = -\xi(\tau_x) - x$$

the first passage time of the level x and the value of the overshoot through this level at the instant of the first passage, the first passage time of the level $-x$ and the value of the overshoot this level at that instant. Integral transforms of the joint distribution of $\{\tau^x, T^x\}$, $\{\tau_x, T_x\}$ for $s > 0$, $\operatorname{Re} p \geq 0$ satisfy the following equalities [Pecherskii and Rogozin (1969) or Zolotarev (1964)]

$$E\,[\,e^{-s\tau^x - pT^x}\,] = \Big(E[\,e^{-p\xi^+(\nu_s)}]\Big)^{-1} E\,[\,e^{-p(\xi^+(\nu_s)-x)};\ \xi^+(\nu_s) > x\,],$$
$$E\,[\,e^{-s\tau_x - pT_x}\,] = \Big(E[\,e^{p\xi^-(\nu_s)}]\Big)^{-1} E\,[\,e^{p(\xi^-(\nu_s)+y)};\ -\xi^-(\nu_s) > x\,],$$

where $\xi^+(t) = \sup_{u \leq t} \xi(u)$, $\xi^-(t) = \inf_{u \leq t} \xi(u)$, ν_s is an exponential variable with parameter $s > 0$, independent of the process, $P[\,\nu_s > t\,] = \exp\{-st\}$, and

$$E[e^{-p\,\xi^{\pm}(\nu_s)}] = \exp\left\{\int_0^\infty \frac{1}{t}\,e^{-st}\,E\,[\,e^{-p\xi(t)} - 1;\ \pm\xi(t) > 0\,]\,dt\right\}, \qquad \pm\operatorname{Re} p \geq 0.$$

3. THE FIRST EXIT FROM THE INTERVAL

Theorem 3.1 *Let* $\{\xi(t);\ t \geq 0\}$, $\xi(0) = 0$ *be a real-valued Lévy process with Laplace exponent (1),* $B > 0$, $y \in [0, B]$, $x = B - y$, *and*

$$\chi(y) = \inf\{\,t > 0:\ y + \xi(t) \notin [0, B]\,\}, \qquad X(y) = (\xi(\chi(y)) - B)\,I_{A^B} + (-\xi(\chi(y)))\,I_{A_0}$$

the instant of the first exit by the process $y + \xi(t)$ *from the interval* $[0, B]$ *and the value of the overshoot through a boundary at the epoch of the exit from the interval by the given process. The Laplace transforms of the joint distribution of* $\{\,\chi(y), X(y)\,\}$ *for* $s > 0$ *satisfy the following formulae*

$$E\,[\,e^{-s\chi(y)};\ X(y) \in du,\ A^B\,] = f_+^s(x, du) + \int_0^\infty f_+^s(x, dv)\,K_+^s(v, du),$$
$$E\,[\,e^{-s\chi(y)};\ X(y) \in du,\ A_0\,] = f_-^s(y, du) + \int_0^\infty f_-^s(y, dv)\,K_-^s(v, du), \tag{2}$$

where

$$f_+^s(x, du) = E\,[\,e^{-s\tau^x};\ T^x \in du\,] - \int_0^\infty E\,[\,e^{-s\tau_y};\ T_y \in dv\,]\,E\,[\,e^{-s\tau^{v+B}};\ T^{v+B} \in du\,],$$
$$f_-^s(y, du) = E\,[\,e^{-s\tau_y};\ T_y \in du\,] - \int_0^\infty E\,[\,e^{-s\tau^x};\ T^x \in dv\,]\,E\,[\,e^{-s\tau_{v+B}};\ T_{v+B} \in du\,];$$

and $K_\pm^s(v, du) = \sum_{n=1}^\infty K_\pm^{(n)}(v, du, s)$, $v \geq 0$ *are the series of the successive iterations;*

$$K_\pm^{(1)}(v, du, s) = K_\pm(v, du, s), \quad K_\pm^{(n+1)}(v, du, s) = \int_0^\infty K_\pm^{(n)}(v, dl, s)\,K_\pm(l, du, s) \tag{3}$$

are the successive iterations $(n \in \mathbb{N} = \{1, 2, \dots\})$ *of the kernels* $K_{\pm}(v, du, s)$*, which are given by the defining equalities*

$$K_{+}(v, du, s) = \int_0^{\infty} E\,[\,e^{-s\tau_{v+B}};\ T_{v+B} \in dl\,]\,E\,[\,e^{-s\tau^{l+B}};\ T^{l+B} \in du\,],$$
$$K_{-}(v, du, s) = \int_0^{\infty} E\,[\,e^{-s\tau^{v+B}};\ T^{v+B} \in dl\,]\,E\,[\,e^{-s\tau_{l+B}};\ T_{l+B} \in du\,]. \qquad (4)$$

4. SUPREMUM, INFIMUM AND THE VALUE OF THE PROCESS

In this section we determine the joint distribution of $\{\inf_{t\le\nu_s} \xi(t),\ \xi(\nu_s),\ \sup_{t\le\nu_s} \xi(t)\}$ for a general Lévy process (i.e. at an exponential time ν_s). Further we will use the following notation: $\xi^-(t) = \inf_{u\le t} \xi(u),\ \xi^+(t) = \sup_{u\le t} \xi(u)$.

Let $\{\xi(t);\ t \ge 0\}$ be a Lévy process with Laplace exponent (1), $x, y \ge 0$, $x + y = B$, $\xi(0) = 0$ and

$$\chi = \inf\{\,t:\ \xi(t) \notin ([-y, x]\,\}, \qquad X = (\xi(\chi) - x)\,I_{A^x} + (-\xi(\chi) - y)\,I_{A_y}$$

the first exit time from the interval $[-y,\ x]$ by the process $\xi(t)$, where $A^x = \{\,\xi(\chi) > x\,\}$, $A_y = \{\,\xi(\chi) < -y\,\}$ are the events on which the exit from $[0, B]$ can occur. Here, unlike in the previous section we shifted the process $y + \xi(t)$ and the interval $[0, B]$ down by y. Note, that due to the space homogeneity of the process the integral transforms of the joint distribution of $\{\,\chi,\ X\,\}$ for the Lévy process with Laplace exponent (1) satisfy formulae (2) of Theorem 2.1.

Observe that,

$$P[\chi > t] = P[-y < \inf_{u\le t} \xi(u),\ \sup_{u\le t} \xi(u) < x].$$

Therefore, we can employ the results of Theorem 2.1 to derive the integral transform of the joint distribution of $\{\,\xi^-(\nu_s),\ \xi(\nu_s),\ \xi^+(\nu_s)\,\}$:

$$Q^s(p) = \int_{-y}^{x} e^{-up}\,P\,[\,-y \le \xi^-(\nu_s),\ \xi(\nu_s) \in du,\ \xi^+(\nu_s) \le x\,] = E\,[\,e^{-p\,\xi(\nu_s)};\ \chi > \nu_s\,]. \qquad (5)$$

Theorem 4.1 *Let* $\{\xi(t); t \ge 0\}$ *be a Lévy process with Laplace exponent (1),* $x,y \ge 0$, $x+y = B$, $\xi(0) = 0$.
The integral transform (5) of the joint distribution of $\{\,\xi^-(\nu_s),\ \xi(\nu_s),\ \xi^+(\nu_s)\,\}$ *satisfies the equality*

$$Q^s(p) = U_p^s(x) - e^{yp} \int_0^{\infty} e^{vp}\,E\,[\,e^{-s\chi};\ X \in dv,\ A_y\,]\,U_p^s(v + B), \qquad \textit{Re}\,p \le 0, \qquad (6)$$

where $(\textit{Re}\,p \le 0)$

$$U_p^s(x) = E\,[\,e^{-p\,\xi(\nu_s)};\ \xi^+(\nu_s) \le x\,] = E[\,e^{-p\xi^-(\nu_s)}]\,E\,[\,e^{-p\,\xi^+(\nu_s)};\ \xi^+(\nu_s) \le x\,]$$

the integral transform of the joint distribution of $\{\, \xi(\nu_s),\ \xi^+(\nu_s)\,\}$ *is given in terms of the Wiener-Hopf factors by the following formula*

$$E\,[\,e^{-p\xi(\nu_s)};\ \xi^+(\nu_s) \le x\,] = E[\,e^{-p\xi^-(\nu_s)}]E\,[\,e^{-p\xi^+(\nu_s)};\ \xi^+(\nu_s) \le x\,], \quad \textit{Re}\, p \le 0;$$

and the integral transforms of the joint distribution of $\{\,\chi,\ X\}$ *are given by (2) of Theorem 2.1.*

For particular classes of Lévy processes formula (6) of Theorem 4.1 takes a simplified form. Let us illustrate it by an example.

Corollary 4.2 *Let* $\{w(t);\ t \ge 0\,\}$ *be a Wiener process with Laplace exponent* $k(p) = \frac{1}{2}\sigma^2 p^2$ *and* $\chi = \inf\{t > 0 :\ w(t) \notin [-y, x]\}$ *be the first exit time from the interval* $[-y,\, x]$*,* $x + y = B$ *by the process* $\{w(t);\ t \ge 0\}$.
Then

1) *the distribution* $\bar{Q}^t(-y, \alpha, \beta, x) \stackrel{\text{def}}{=} P[-y \le \inf\limits_{u\le t} w(u),\ w(t) \in (\alpha,\, \beta),\ \sup\limits_{u\le t} w(u) \le x]$ *is given by formula:*

$$\begin{aligned} &P[-y \le \inf_{u\le t} w(u),\ w(t) \in (\alpha,\, \beta),\ \sup_{u\le t} w(u) \le x] \\ &= \frac{4}{\pi}\sum_{\nu=1}^{\infty}\frac{1}{\nu}\ \exp\left(-\frac{1}{2}t(\pi\nu\sigma/B)^2\right)\ \sin\left(\frac{x}{B}\pi\nu\right)\ \sin\left(\frac{2x-\alpha-\beta}{2B}\pi\nu\right)\ \sin\left(\frac{\beta-\alpha}{2B}\pi\nu\right), \\ &P[\chi > t] = \frac{4}{\pi}\sum_{\nu=0}^{\infty}\frac{1}{2\nu+1}\ \exp\left(-\frac{1}{2}t(\pi(2\nu+1)\sigma/B)^2\right)\sin\left(\frac{x}{B}(2\nu+1)\pi\right), \end{aligned}$$

2) *the moments of the first exit time* χ *are of the following form*

$$E[\chi] = \frac{1}{\sigma^2}xy, \qquad E[\chi^2] = \frac{1}{3\sigma^4}xy(x^2+3xy+y^2), \qquad Var[\chi] = \frac{1}{3\sigma^4}xy(x^2+y^2),$$

in particular, when $x = y$

$$E[\chi^n] = \frac{1}{(2n-1)!!}\left(\frac{x}{\sigma}\right)^{2n} E_n, \qquad\qquad n > 0,$$

where E_n*,* $n > 0$ *are the Euler numbers,* $E_1 = 1$*,* $E_2 = 5$*, ...*

3) *the probability* $\bar{Q}^t(-y, \alpha, \beta, x) = P[-y \le \inf\limits_{u\le t} w(u),\ w(t) \in (\alpha,\, \beta),\ \sup\limits_{u\le t} w(u) \le x]$ *satisfies the formula*

$$\bar{Q}^t(-y, \alpha, \beta, x) = \frac{1}{\sigma\sqrt{2\pi t}}\int_{\alpha}^{\beta}\left(\sum_{k=-\infty}^{\infty} e^{-(2Bk+u)^2/2\sigma^2 t} - \sum_{k=-\infty}^{\infty} e^{-(2Bk+2x-u)^2/2\sigma^2 t}\right) du.$$

5. NUMBER OF INTERSECTIONS

The objective of this section is to determine the joint distribution of the number of upward and downward intersections of the interval $[-y, x]$ by the Lévy process $\{\xi(t);\ t \geq\}\ 0$ with Laplace exponent (1). Assuming that $\xi(0) = -(v + y),\ v > 0$ we denote by $i_v = \inf\{\, t :\ \xi(t) > x\,\}$ the instant of the first upward intersection of the interval $(-y,\ x)$. Assuming that $\xi(0) = v + x, v > 0$ we denote by $i^v = \inf\{\, t :\ \ \xi(t) < -y\,\}$ the instant of the first downward intersection of the interval $[-y,\ x]$. Now let $\xi(0) = 0$ and introduce:

- α_t^+ i.e. the number of the upward intersections of the interval $[-y,\ x]$ up to the instant t;
- α_t^- i.e. the number of the downwards intersections of the interval $[-y,\ x]$ up to the instant t.

Theorem 5.1 *Let* $\{\xi(t);\ t \geq 0\}$, $\xi(0) = 0$ *be a Lévy process with Laplace exponent (1),* $B > 0$, $x \in [0, B]$, $y = B - x$.
Then the joint distribution of the number of upward and downward intersections $\{\ \alpha_{\nu_s}^+, \alpha_{\nu_s}^-\ \}$ *of the interval* $[-y, x]$ *for* $n \in \mathbb{N} \cup \{0\}$ *satisfies the following equalities*

$$P\,[\,\alpha_{\nu_s}^+ = n,\ \alpha_{\nu_s}^- = n+1\,]$$
$$= \int\limits_0^\infty E[\,e^{-s\chi};\ X \in dv,\ A^x\,] \int\limits_0^\infty K_+^{(n)}(v, du, s) \int\limits_0^\infty E[\,e^{-s\tau_{u+B}};\ T_{u+B} \in dl\,]\ \Big(1 - E[e^{-s\tau^{l+B}}]\Big)\,;$$

$$P\,[\,\alpha_{\nu_s}^+ = n+1,\ \alpha_{\nu_s}^- = n\,]$$
$$= \int\limits_0^\infty E[\,e^{-s\chi};\ X \in dv,\ A_y\,] \int\limits_0^\infty K_-^{(n)}(v, du, s) \int\limits_0^\infty E[\,e^{-s\tau^{u+B}};\ T^{u+B} \in dl\,]\ \big(1 - E[e^{-s\tau_{l+B}}]\big)\,,$$

$$P\,[\,\alpha_{\nu_s}^+ = n = \alpha_{\nu_s}^-\,] = I_{\{n=0\}} -$$
$$- I_{\{n=0\}} \left(\int\limits_0^\infty E[\,e^{-s\chi};\ X \in dv,\ A^x\,] E[\,e^{-s\tau_{v+B}}] + \int\limits_0^\infty E[\,e^{-s\chi};\ X \in dv,\ A_y\,] E\,[e^{-s\tau^{v+B}}] \right)$$
$$+ I_{\{n \in \mathbb{N}\}} \int_0^\infty E[\,e^{-s\chi};\ X \in dv,\ A^x\,] \int_0^\infty K_+^{(n)}(v, du, s)\ \big(1 - E[e^{-s\tau_{u+B}}]\big)$$
$$+ I_{\{n \in \mathbb{N}\}} \int_0^\infty E[\,e^{-s\chi};\ X \in dv,\ A_y\,] \int_0^\infty K_-^{(n)}(v, du, s)\ \Big(1 - E[e^{-s\tau^{u+B}}]\Big)\,;$$

where $K_\pm^{(0)}(v, du, s) \stackrel{\text{def}}{=} \delta(v - u)\, du$, *and the functions* $E\,[\,e^{-s\chi};\ X \in dv,\ A^x\,]$, $E\,[\,e^{-s\chi};\ X \in dv,\ A_y\,]$, *and the successive iterations* $K_\pm^{(n)}(v, du, s)$,$n \in \mathbb{N}$ *of the kernels* $K_\pm(v, du, s)$ *are given by the formulae (3) of Theorem 2.1.*

As an example, consider a Wiener process $\{w(t);\ t\geq 0\}$ with Laplace exponent $k(p)=\frac{1}{2}\sigma^2p^2$, i.e. $P[w(t)\in(a,b)]=\frac{1}{\sigma\sqrt{2\pi t}}\int\limits_a^b e^{-u^2/2\sigma^2 t}du$.
Then

$$P[\alpha_t^+=n+1,\ \alpha_t^-=n]=2\sum_{k=2(n+1)}^{\infty}(-1)^kP[w(t)\in(-x+kB,x+kB)];$$

$$P[\alpha_t^-=n+1,\ \alpha_t^+=n]=2\sum_{k=2(n+1)}^{\infty}(-1)^kP[w(t)\in(-y+kB,y+kB)];$$

$$\begin{aligned}&P[\alpha_t^+=\alpha_t^-=n]\\&=2(1-\delta_{n0})\sum_{k=2(n+1)}^{\infty}(-1)^kP[w(t)\in(-x+kB,x+kB)]\\&\quad+2(1-\delta_{n0})\sum_{k=2(n+1)}^{\infty}(-1)^kP[w(t)\in(-y+kB,y+kB)]\\&\quad+\delta_{n0}\left(1-2\sum_{k=0}^{\infty}(-1)^k\left\{P[w(t)>x+(k+1)B]+P[w(t)>y+(k+1)B]\right\}\right).\end{aligned}$$

6. FIRST ENTRY TIME

In this section we determine the integral transforms of the joint distribution of the first entry time into the interval $[0,B]$ and the value of the Lévy process at this instant. Observe, that the first entry of the interval (after leaving it) can take place either from above (position of the process: $v+B$), or from below the interval ($-v$), or from the starting point ($\xi(0)=0$).

Theorem 6.1 *Let $\xi(t)\in\mathbb{R}$, $t\geq 0$, $\xi(0)=0$ be a Lévy process with Laplace exponent (1), $B>0$, $\chi(y)\stackrel{\text{def}}{=}0$, for $y\notin[0,B]$, and*

$$\overline{\chi}(y)=\inf\{\,t>\chi(y):\ y+\xi(t)\in[0,B]\,\},\qquad \overline{X}(y)=y+\xi(\overline{\chi}(y))\in[0,B],\quad y\in\mathbb{R}$$

the first entry time of the process $y+\xi(t)$ into $[0,B]$ and the value of the process $y+\xi(t)$ at the epoch of the entry.
Then the integral transforms of the joint distribution of $\{\overline{\chi}(y),\overline{X}(y)\}$, $y\in\mathbb{R}$ for $s>0$ satisfy the formulae

$$\begin{aligned}b^v(du,s)&=E\,[\,e^{-s\overline{\chi}(v+B)};\ \overline{X}(v+B)\in du\,]\\&=\int_0^\infty Q_+^s(v,dl)\,E[\,e^{-s\tau_l};\ B-T_l\in du\,]\\&\quad+\int_0^\infty Q_+^s(v,dl)\int_0^\infty E[\,e^{-s\tau_l};\ T_l-B\in d\nu\,]\,E[\,e^{-s\tau^\nu};\ T^\nu\in du\,],\qquad v>0\end{aligned}$$

$$\begin{aligned}
b_v(du,s) &= E\,[\,e^{-s\overline{\chi}(-v)};\ \overline{X}(-v)\in du\,]\\
&= \int_0^\infty Q_-^s(v,dl)\,E[\,e^{-s\tau^l};\ T^l\in du\,]\\
&\quad+\int_0^\infty Q_-^s(v,dl)\int_0^\infty E[\,e^{-s\tau^l};\ T^l-B\in d\nu\,]\,E[\,e^{-s\tau_\nu};\ B-T_\nu\in du\,],\quad v>0,\\
b(y,du,s) &= E\,[\,e^{-s\overline{\chi}(y)};\ \overline{X}(y)\in du\,]\\
&= \int_0^\infty E\,[\,e^{-s\chi(y)};\ X(y)\in dv,\ A^B]\,b^v(du,s)\\
&\quad+\int_0^\infty E\,[\,e^{-s\chi(y)};\ X(y)\in dv,\ A_0]\,b_v(du,s), \qquad y\in[0,B],
\end{aligned}$$

where $\delta(x)$, $x\in\mathbb{R}$ is the delta function,

$$Q_\pm^s(v,du)=\delta(v-u)\,du+\sum_{n\in\mathbb{N}}Q_\pm^{(n)}(v,du,s),\qquad v>0,$$

are the series of the successive iterations $Q_\pm^{(n)}(v,du,s)$, and $n\in\mathbb{N}$,

$$Q_\pm^{(1)}(v,du,s)=Q_\pm(v,du,s),\qquad Q_\pm^{(n+1)}(v,du,s)=\int_0^\infty Q_\pm^{(n)}(v,dl,s)Q_\pm(l,du,s),$$

are successive iterations of the kernels $Q_\pm(v,du,s)$, which are defined by the following formulae

$$\begin{aligned}
Q_+(v,du,s)&=\int_0^\infty E[e^{-s\tau_v};\ T_v-B\in dl]\,E[e^{-s\tau^l};\ T^l-B\in du],\\
Q_-(v,du,s)&=\int_0^\infty E[e^{-s\tau^v};\ T^v-B\in dl]\,E[e^{-s\tau_l};\ T_l-B\in du].
\end{aligned}$$

Remark 6.1 The proofs of all stated theorems can be found in Kadankov and Kadankova (2005). Examples of the stated results for particular classes of Lévy process are also given in Kadankova and Veraverbeke (2006), Kadankova (2003).

Remark 6.2 Although we obtained rather sophisticated expressions for integral transforms of the two-boundary characteristics of the Lévy process, we hope that the results presented can be of use in further investigation of Lévy processes and give some fruitful ideas for further applications in finance.

References

S. Asmussen and J. Rosinski. Approximations of small jumps of Lévy processes with a view towards simulation. *Journal of Applied Probability*, 38(2):482–493., 2001.

S. Asmussen, F. Avram, and M.R. Pistorius. Russian and American put options under exponential phase-type Lévy models. *Stochastic Processes and their Applications*, 109:79–111, 2004.

S. Asmussen, D. Madan, and M.R. Pistorius. Pricing equity default swaps under the CGMY Lévy model. *Submitted*, 2005.

F. Avram, A.E. Kyprianou, and M.R. Pistorius. Exit problems for spectrally negative Lévy processes and applications to Russian, American and Canadized options. Technical report, Utrecht University, 2002.

J. Bertoin. *Lévy processes*. Cambridge University Press, Cambridge, 1996.

J. Bertoin. Exponential decay and ergodicity of a completely assymetric Lévy process in a finite interval. *Annals of Applied Probability*, 1:156–169, 1997.

S. Boyarchenko and S. Levendorskii. Barrier options and touch-and-out options under regular Lévy processes of exponential type. *Annals of Probability*, 12(4):1261–1298, 2002.

H. Geman. Pure jump Lévy processes for asset price modelling. *Journal of Banking and Finance*, 26(7):1297–1316, 2002.

V.F. Kadankov and T. Kadankova. On the distribution of the first exit time from an interval and the value of overshoot through the boundaries for processes with independent increments and random walks. *Ukrainian Mathematical Journal*, 10(57):1359–1384, 2005.

T. Kadankova. On the distribution of the number of the intersections of a fixed interval by the semi-continuous process with independent increments. *Theory of Stochastic Processes*, 1-2: 73–81, 2003.

T. Kadankova and N. Veraverbeke. On several two-boundary problems for a certain class of Lévy processes. Submitted, 2006.

A.E. Kyprianou and M.R. Pistorius. Perpetual options and Canadization through fluctuation theory. *Annals of Applied Probability*, 13(3):1077–1098, 2003.

E.A. Pecherskii and B.A. Rogozin. On joint distributions of random variables associated with fluctuations of a process with independent increments. *Theory of Probability and its Applications*, 14:410–423, 1969.

M.R. Pistorius. A potential theoretical review of some exit problems of spectrally negative Lévy processes. *Séminaire de Probabilités*, 38:30–41, 2004.

T. Rydberg. The normal inverse Gaussian Lévy process: simulation and approximation. *Communications in Statistics: Stochastic models*, 13:887–910, 1997.

K.-I. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

W. Schoutens. The Meixner process: Theory and applications in finance. Technical report, EURANDOM, Eindhoven, 2001.

W. Schoutens. *Lévy processes in finance*. Wiley, Chichester, England, 2003.

W. Schoutens. Exotic options under Lévy models: an overview. *Journal of Computational and Applied Mathematics*, 189:526–538, 2004.

A.V. Skorokhod. *Theory of Random Processes*. Boston Spa, Yorkshire, England, 1971.

V. M. Zolotarev. The first passage time of a level and the behaviour at infinity for a class of processes with independent increments. *Theory of Probability and its Applications*, 9:653–664, 1964.

ECONOMIC CAPITAL ALLOCATION UNDER LIQUIDITY CONSTRAINTS

Fernando Mierzejewski

Faculty of Economy and Applied Economy, Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium
Email: Fernando.Mierzejewski@wis.kuleuven.be

Abstract

Since the capital structure affects the performance of financial institutions confronted to liquidity constraints, the *Economic Capital* is determined by the maximisation of value. Allowing economic decisions to be characterised by a *distorted* probability distribution — so assessing the attitude towards risk as well as information and knowledge — the optimal surplus is expressed as a *Value-at-Risk* — as recommended by the Basel Committee. Thus, demanding more capital than regulatory requirements accounts for different expectations about risks. The optimal surplus is allocated to the lines of business of a conglomerate according to the borne risk and the type of divisional managers. Full allocation is assured and no covariances are required. Further, a mechanism is provided, which allows for the distribution of equity in a decentralised organisation.

1. INTRODUCTION

In a seminal paper, Modigliani and Miller (1959) claimed that in perfect markets the capital structure of financial institutions does not matter for at any time it is possible to raise or release funds if required. Accordingly, the optimal plan when the objective is maximising value, is to attract as much debt as possible. Since this fact is not observed in practice, Modigliani and Miller gave several explanations in subsequent papers, even questioning the skills of decision makers, as in Miller (1998). However, averse-to-risk customers are sensible to fluctuations and then the performance of intermediaries depends on providing guarantees that assumed liabilities are default-free, see Merton (1997). This situation leads manager's decisions to be determined also by risk aversion — as long as their reputation depends on performance.

Usual practices to protect against default risk are *hedging*, *re-insuring* and *capital cushions*. By *Economic* or *Risk Capital* we mean an amount of money invested in non risky assets that serves as a buffer in order to prevent insolvency. Since a price has to be paid for raising capital, there is a level of surplus which properly combines the two conflicting objectives: maximisation of shareholder's value and minimisation of default risk. Within a multibusiness environment, the problem of allocation arises due to the gain acquired — through diversification — when merging the activities of the firm. Such benefit should be distributed fairly among the subsidiaries — i.e. according to the risk borne. In this context, many of the allocation principles present in the literature are based on covariances. Full allocation is also considered as a desirable property — for the aggregate surplus maintained by divisions should be equal to the level regarded as appropriate for the conglomerate[1].

In Merton and Perold (1993) a capital allocation principle is developed based on the incremental risk of subsidiaries, which is obtained by subtracting the capital required after suppressing a line of business to the surplus demanded by the whole portfolio. Then the sum of individual surplus might be lower than the capital hired by the conglomerate — the difference is explained by the gains in efficiency due to the knowledge of divisional managers. On this basis, Merton and Perold argue that it is inappropriate to full allocate the capital — for doing so incentives may be distorted. Myers and Read Jr. (2001) consider instead the marginal capital requirement, defined as the marginal change in the total surplus in response to a small increment in the equity demanded by a certain line of business. They prove that full allocation is guaranteed by this principle, provided that some conditions on the valuation function of capital are satisfied.

Stoughton and Zechner (1999) propose a model to deal with firms that are not able to continuously raise funds — see also Froot et al. (1993) and Jensen (1986). Thus, equity is distributed in order to maximise the *Economic Value Added (EVA)* by the lines of business, and capital allocation is justified as a mechanism that stimulates the exchange of information inside the institution. In the process, the attitude towards risk is considered, which is supposed to depend on the ability to apply and transfer skills — as well as the effort expended to accumulate information. Thus, an optimal mechanism is advanced based on the internal price of capital. Distortions are allowed in the form of under and overinvestment.

In the following, an allocation principle is proposed which, instead of accounting for stochastic dependencies, focuses on agency costs due to discrepancies in the expectations kept by central and divisional managers. Actually, the case of perfect correlation is considered — when no diversification is possible — in this way modelling the situation when the failure in any division may damage the credit quality of the whole conglomerate. *Section 2* is devoted to the determination of the optimal amount of economic capital. The attitude towards risk is determined by a single — functional — parameter, which in imperfect markets accounts for differences in expectations among decision makers. Thus, the demanded surplus depends on the risk profile — or the informational type — of decision makers, as well as on the risk involved. The problem of capital allocation within a multibusiness setting is addressed in *Section 3*. A *centralised* solution is obtained depending on individual exposures. In *Section 4* the *stand alone* allocation is attained by letting subsidiaries to act on their own. Finally, the problem of agency costs is addressed by establishing an *optimal contract*. When the types are not accessible — a situation most probably found in practice — a mechanism can be designed by fixing the cost of raising capital inside the conglomerate. In this way, subsidiaries are forced to reveal their type. *Section 5* concludes.

[1] See Albrecht (2004), Hallerbach (2003) and Saita (2004).

2. ECONOMIC CAPITAL AS THE OPTIMAL LEVEL OF SURPLUS

Consider a financial institution holding assets and liabilities for total market values of A and L respectively. The net — random — loss suffered each period is then given by $X = L - A$. Merton (1977) defines the fair price of insuring liabilities — at any time before the maturity date — as the present value of the liability claim less the value of a put option on assets with strike price equal to the value of liabilities[2]. In the same way, it follows that shareholders are the owners of a call option on the portfolio of assets whose exercise price is the value of liabilities. From the *Put-Call Parity Theorem*, the following relation must hold:

$$A = C(A, L) + Le^{-r_0 T} - P(A, L).$$

Thus, though both the market value of assets and equity are functions of leverage, by the *Put-Call Parity* their sum is independent of it. Hence, the market value of the firm, i.e. the market value of the portfolio of assets A, is independent of the capital structure, as stated in the Modigliani Miller proposition, see Miller (1998). However, this reasoning holds true in perfect markets, i.e. when no restrictions are to be found when borrowing and lending. Moreover, the hedged portfolio remains non risky only a short period of time ahead, assuming that during a short period of time market conditions remain unchanged. Thus, continuous rebalancing is needed. Under these conditions, the conglomerate will be indifferent between hedging and reinsurance. But decision makers confronted to liquidity constraints might be interested in replacing — or complementing — their hedging strategy.

By now, assume that central managers know the distribution function of losses F_X and that funds may be hired at the interest rate r_k, with $r_k \geq r_0$, where r_0 denotes the risk free interest rate. Decisions are affected by the *net cost of capital* $\eta_k = r_k - r_0$. Moreover, notice the firm simultaneously acts in two markets. So whenever a loss occurs cash is demanded to avoid default, while in the case a gain is obtained, the surplus can be used to buy more assets or to pay liabilities. Assuming that investors keep different expectations about risk — as long as they own different information, knowledge, social contacts and capabilities — and denoting respectively by φ and β the types for lending and borrowing, *corporate EVA* is given by:

$$\text{EVA} \; = \; \text{E}_{\varphi}\left[(X+k)_{-}\right] - \text{E}_{\beta}\left[(X-k)_{+}\right] - \eta_k k.$$

The term $\text{E}_{\varphi}\left[(X+k)_{-}\right]$ denotes the value of the firm when the portfolio is solvent, i.e. when $X < -k$, which is diminished by raising the level of surplus. On the other hand, the term $\text{E}_{\beta}\left[(X-k)_{+}\right]$ represents the *cost of bankruptcy* — or more properly, the cost of *assuming* bankruptcy. Demanding more capital leads to a reduction of the burden of default. Thus, financial intermediaries are able to create value to shareholders as long as the cost of insuring the aggregate exposure — which can be related to the credit quality, as perceived by lenders — plus the cost of raising capital is less than expected gains. Notice how crucial is the role played by the differences in expectations and the symmetry of risks. Under homogeneous expectations and symmetric risks, keeping a surplus produces a total loss and so no capital should be hired — the value of the firm in

[2]Whenever $A \geq L$ the firm can afford the debt, but when $A < L$ the guarantor suffers a loss equal to $L - A$. Consequently, the guarantor's claim equals $\min(A - L, 0)$ which is identical to that of a put option — where the promised payment L corresponds to the exercise price and the value of assets corresponds to the common stock's price. See also Cummins and Sommer (1996).

this case is zero, which is a reasonable claim in a competitive setting. In this way, the result of the Modigliani and Miller proposition is obtained, see Stiglitz (1972).

The *Wang's risk principle* allows for a characterisation of the mathematical expectation with respect to a *distorted* probability distribution, which is obtained by applying a *distortion* transformation — i.e. a continuous, strictly increasing function, defined on the unit interval $\varphi : [0,1] \to [0,1]$ such that $\varphi(0) = 0$ and $\varphi(1) = 1$ — to the *decumulative distribution function* $S_X(x) = 1 - F_X(x) = \mathrm{P}[X > x]$ in the following way[3]:

$$\mathrm{E}_\varphi[X] = \int x dF_{\varphi,X}(x) = \int [1 - F_{\varphi,X}(x)]\, dx = \int \varphi(S_X(x))dx.$$

The traditional expectation operator is obtained when the *neutral distortion*, equal to the identity operator $\varphi(x) = x$, $\forall\, x$, is introduced. Further, Wang and Young (1998) state the properties:

$$\begin{aligned} \varphi \text{ concave } &\Rightarrow\ \varphi(y) \geq y \ \ \forall y \in [0,1] \ \Rightarrow\ \mathrm{E}_\varphi[X] \geq \mathrm{E}[X] \\ \varphi \text{ convex } &\Rightarrow\ \varphi(y) \leq y \ \ \forall y \in [0,1] \ \Rightarrow\ \mathrm{E}_\varphi[X] \leq \mathrm{E}[X]. \end{aligned}$$

Therefore, *concave* distortion functions characterise the decisions of *averse-to-risk* investors — who overestimates risks — and *convex* distortions the behaviour of *risk lovers* — who underestimates risks. Moreover, applying a *Taylor series* around zero leads to:

$$\mathrm{E}_\varphi\left[(X+k)_-\right] \approx \mathrm{E}_\varphi[X_-] + \left[\frac{\partial\, \mathrm{E}_\varphi\left[(X+k)_-\right]}{\partial k}(k=0)\right] \cdot k.$$

Let us accordingly define:

$$r_{\varphi,X} := -\frac{\partial\, \mathrm{E}_\varphi\left[(X+k)_-\right]}{\partial k}(k=0) = F_{\varphi,X}(0).$$

The coefficient $r_{\varphi,X}$ corresponds to a *premium for solvency* — specifically, it expresses the marginal reduction of the insured return when hiring an additional unit of equity. When the risk accumulates more probability in gains — remember the variable X represents an aggregated loss — a higher premium has to be paid. On this basis, the level of *Economic Capital* is determined in order to maximise corporate EVA[4]:

$$\mathbf{Max}_k\ \mathrm{E}_\varphi[X_-] - \mathrm{E}_\beta\left[(X-k)_+\right] - (r_{\varphi,X} + \eta_k)k.$$

Applying *Lagrange optimisation* yields the first order condition:

$$-\frac{\partial}{\partial k}\mathrm{E}_\beta\left[(X-k)_+\right] - (r_{\varphi,X} + \eta_k) = S_{\beta,X}(k^*) - (r_{\varphi,X} + \eta_k) = 0.$$

[3]The distorted probability principle is extended to real-valued random variables as, see Wang et al. (1997): $\mathrm{E}_\varphi[X] = \int_{-\infty}^0 [\varphi(S_X(t)) - 1]\, dt + \int_0^\infty \varphi(S_X(t))dt$. Hence, after performing a change of variables, we can write: $\mathrm{E}_\varphi[X] + \mathrm{E}_\varphi[X_-] = \mathrm{E}_\varphi[X_+]$. The right-hand-side of the equation shows the price of a portfolio containing an insured version of the asset, while the left-hand-side shows the price of a fund containing the asset and a guarantee to pay the loss incurred by X. Both portfolios have the same value at the end of period, and hence both should be assigned the same market price. Therefore the condition is consistent with the no-arbitrage principle.

[4]A raising principle is presented in this fashion by Dhaene et al. (2003) though they propose to minimise the total capital cost, see also Goovaerts et al. (2005), Laeven and Goovaerts (2004) and Froot et al. (1993).

Hence, the firm attracts debt until the marginal benefit equals the total cost of capital and the optimal level of surplus is given by:

$$k^* = F^{-1}_{\beta,X}(1 - r_{\varphi,X} - \eta_k) = S^{-1}_{\beta,X}(r_{\varphi,X} + \eta_k) = S^{-1}_X\left(\beta^{-1}(r_{\varphi,X} + \eta_k)\right).$$

The term $(r_{\varphi,X} + \eta_k)$ accounts for the *total cost* of holding an additional unit of capital. When this index is high — i.e. when a high premium is asked for solvency or a high cost is confronted when attracting liabilities — less equity is provided. The contrary occurs when the total cost is low — i.e. when the premium for solvency or the price of capital is low. Whenever $(r_{\varphi,X} + \eta_k) \geq 1$ and $(r_{\varphi,X} + \eta_k) \leq 0$, the minimum and the maximum level of cash are preferred respectively. There is an additional motivation to demand as much surplus as possible in the later case, for the deterioration in the credit quality of the firm might raise the net cost η_k. Moreover, averse-to-risk investors, for whom the distortion function is concave, so that $\varphi^{-1}(\eta) < \eta$, underestimate the price of equity.

The optimal amount of capital — or the *Economic Capital* — is thus expressed as a *Value-at-Risk* under a transformed probability measure. This criterion coincides with the capital requirement established by the Basel Capital Accord[5]. Accordingly, the *Regulatory Capital* is obtained by applying the *neutral distortion* and introducing a *level of confidence* α — in this way implicitly determining the premium for solvency as well as the cost of capital by letting $\alpha = r_{\varphi,X} + \eta_k$. Typically, $\alpha = 5\%$ or $\alpha = 1\%$. Since the same confidence level is asked for every company, the most efficient — which are asked a higher premium for they hold better investments — are forced to keep more surplus than the optimal level. This loss in efficiency makes sense from the perspective of the regulator, as long as the social losses produced because of the simultaneous default of many firms in the industry might be huge — by affecting the economic activity and the aggregate demand. But on the other hand, the minimum level required for the intermediaries that perform badly might be underestimated.

3. OPTIMAL ALLOCATION OF ECONOMIC CAPITAL AMONG LINES OF BUSINESS

In order to hold the viewpoint of central managers, or a regulatory authority, confronting a multi-business environment, let us suppose that X denotes the aggregate loss of a financial conglomerate consisting of $n \in N$ subsidiaries, or lines of business, such that X equals the sum of individual risks:

$$X = X_1 + \cdots + X_n.$$

Marginal distributions $(F_1, \ldots, F_n)$ are assumed to be known and since a failure in any division may damage the reputation of the whole conglomerate, the *comonotonic dependence structure* is considered[6]. When capital decisions are centralised, the cost of the guarantee can be diminished by merging the individual losses[7], for in this way funds can be assigned only to insolvent divisions —

[5]See Basel Committee on Banking Supervision 1996 and Basel Committee on Banking Supervision 2004.

[6]Comonotonicity characterises an extreme case of dependence, when no benefit can be obtained from diversification, see Dhaene et al. (2002).

[7]Mathematically, this result is sustained by the fact that the distorted probability principle preserves the first stochastic order, defined by $X \leq Y \;\Leftrightarrow\; S_X(t) \leq S_Y(t), \forall t$. Therefore, $\mathrm{E}_\varphi\left[(X-k)_+\right] \leq \mathrm{E}_\varphi\left[\sum_{i=1}^n (X_i - k_i)_+\right]$ when $\sum_{i=1}^n k_i = k$. See Goovaerts et al. (2005) and Laeven and Goovaerts (2004).

and no idle surplus is maintained. Accordingly, let us establish an allocation principle based on the minimisation of the sum of exposures — the value of the firm is already maximised by choosing the level k^* as the total surplus kept by the conglomerate:

$$\textbf{Min}_{k_i}\ \mathrm{E}_\varphi\left[\sum_{i=1}^{n}(X_i-k_i)_+\right]$$
$$\textit{subject to}\ \sum_{i=1}^{n}k_i=k^*.$$

Therefore, a diversification effect exists, but it depends on liquidity constraints — and not on covariances. The only condition imposed is full allocation — as long as capital decisions on business units are taken by central managers, no other concerns are needed. For the *Lagrange multiplier* γ the first order conditions are the following:

$$\frac{\partial}{\partial k_i}\mathrm{E}_\varphi\left[\sum_{i=1}^{n}(X_i-k_i)_+\right]+\gamma\ =\ -S_{\varphi,X_i}\left(k_i^*\right)+\gamma\ =\ 0\quad \forall\, i=1,\ldots,n$$
$$\sum_{i=1}^{n}k_i^*=k^*.$$

Let us denote by F_{X^c} the probability distribution of the *comonotonic sum* $X^c=X_1^c+\cdots+X_n^c$, where $(X_1^c,\ldots,X_n^c)$ represents the *comonotonic random vector* with same marginal distributions as $(X_1,\ldots,X_n)$. Since the inverse distribution of the comonotonic sum is given by the sum of the inverse marginal distributions, see Dhaene et al (2002), we get that γ is determined such that:

$$F^{-1}_{\varphi,X^c}(1-\gamma)\ =\ \sum_{i=1}^{n}F^{-1}_{\varphi,X_i}(1-\gamma)\ =\ \sum_{i=1}^{n}k_i^*\ =\ k^*.$$

Thus, the optimal risk capitals allocated to the business units are given by:

$$k_i^*\ =\ F^{-1}_{\varphi,X_i}\left(F_{\varphi,X^c}\left(k^*\right)\right)\quad \forall\, i=1,\ldots,n.$$

These levels of equity determine the *centralised solution* — for both the raising and the allocation principles have been established according to the risk attitude and knowledge of central managers.

4. OPTIMAL DECENTRALISED MECHANISM

Full allocation suffices for centralised organisations. But divisions are run by managers who access better information about investment opportunities, a situation that leads shareholders to incur in agency costs, see Jensen (1986). So let us consider subsidiaries as separate units that maximise value but do not assume the reduction of the insured return — and hence do not internalise the premium for solvency in decision making. By putting the burden of bankruptcy on their shoulders, central managers attain a gain due to the diversification of the liquidity constraint, as stated in *Section 3*. Accordingly, as long as subsidiaries hire capital from central management at the net internal cost η, *divisional EVA* is defined in the following way:

$$\text{EVA} \;=\; \mathrm{E}_{\varphi_i}\left[(X_i)_-\right] - \mathrm{E}_{\varphi_i}\left[(X_i - k_i)_+\right] - \eta k.$$

Therefore, divisions maximise value by minimising the total loss $\mathrm{E}_{\varphi_i}\left[(X_i - k_i)_+\right] + \eta k$. After the first order condition, the *stand alone risk capital* is determined by:

$$k_i(\eta) = F^{-1}_{\varphi_i, X_i}(1-\eta) \quad \forall\, i = 1, \ldots, n.$$

By means of the net cost η, the capital decisions of subsidiaries may be distorted — forcing them to internalise bankruptcy according to the interest of the conglomerate. So in order to encourage averse-to-risk managers to raise less capital, its cost might be overcharged. A return over the market rate r_k should be assigned in this situation such that $\eta > \eta_k$. On the contrary, risk lovers might be subsidised so that $\eta < \eta_k$ — for giving them incentives to hire more capital. The optimal levels of economic capital and internal cost are simultaneously determined by introducing the following allocation principle, see Diamond and Verrecchia (1982):

$$\begin{aligned} &\mathbf{Max}_{k,\eta}\;\; \mathrm{E}_{\varphi}\left[\, X_- \right] - \mathrm{E}_{\varphi}\left[\, (X - k)_+ \right] - (r_{\varphi,X} + \eta_k) \cdot k \\ &\textit{subject to} \quad k_i = k_i(\eta) \;\; \textit{and} \;\; \sum_{i=1}^{n} k_i = k. \end{aligned}$$

Applying *Lagrange optimisation* leads the solution to be characterised by:

$$S_{\varphi,X}\left(k^*\right) = r_{\varphi,X} + \eta_k \quad \text{and} \quad \sum_{i=1}^{n} k_i^* = k^*.$$

Hence, the same optimal surplus of *Section 2* is obtained for the conglomerate, while the internal cost of capital is determined such that full allocation is assured:

$$\sum_{i=1}^{n} F^{-1}_{\varphi_i, X_i}\left(1 - \eta^*\right) = k^*.$$

Therefore, if $F_{\varphi_1, \ldots,\, \varphi_n, X^c} = \left(\sum_{i=1}^{n} F^{-1}_{\varphi_i, X_i}\right)^{-1}$ denotes the distribution function of the comonotonic sum when marginal distributions are given by $(F_{\varphi_1, X_1}, \ldots, F_{\varphi_n, X_n})$, then the optimal level of the net internal cost of capital is given by:

$$\eta^* \;=\; 1 - F_{\varphi_1, \ldots,\, \varphi_n, X^c}\left(k^*\right).$$

In this way, a *decentralised* allocation is determined — the same benefit as under the *centralised* prescription is obtained and so no efficiency is lost. When the types of subsidiaries are not observable, central managers may calibrate their estimations by comparing the preferred amounts of equity with the optimal levels k_i^*. Therefore, by letting divisional managers to act independently they are forced to reveal their type. We can then say the proposed mechanism provides a basis to measure the disagreement between central management and business units.

5. CONCLUSIONS

According to the Modigliani and Miller (1959) proposition, the capital structure of a financial institution does not affect its value for it is always possible to raise or release funds in the market. However, this is not a suitable assumption for imperfect markets. Actually, after Merton (1997), the level of surplus matters for averse-to-risk lenders who are sensible to the possibility of bankruptcy of the borrower. Accordingly, the decisions of managers, whose reputation depends on performance, are also affected by risk aversion. In this context, the value of the firm depends on its capital structure, and the *Economic Capital* is defined such that the *Economic Value Added (EVA)* is maximised.

The *Wang's principle*, see Wang et al. (1997), allows expressing the *cost of bankruptcy* as an expectation with respect to a *distorted probability distribution*. The — functional — distortion type simultaneously accounts for risk attitude and knowledge, and investors are supposed to maintain different expectations — an approach already adopted by Stiglitz (1972). The optimal level of surplus is then a function of the total cost of equity — defined as the premium for solvency plus the net capital cost — as well as the risk involved, and since no restrictions are imposed on the distribution functions of returns, the model is suitable both to financial and insurance applications. Thus, decision makers internalise the price of equity, though it is underestimated by risk averse investors who apply a concave transformation to the probability distribution, and consequently demand more capital.

Specifically, the *Economic Capital* is expressed as a *Value-at-Risk* under a distorted probability measure, at the time the *Regulatory Capital* is obtained by applying no distortion and fixing a confidence level α — which in this way plays the same role as the total equity cost and hence in the model both coefficients are given the same meaning. Capital decisions over the minimum regulatory requirement are then explained by risk aversion — for payments are overestimated in this case. However, risk lover investors may overestimate exposures as well, as long as the type also accounts for information and knowledge. In this context, the excess of surplus induces a gain in efficiency, and not the opposite.

A *centralised* allocation of equity is determined by maximising *corporate EVA* and minimising *bankruptcy costs* according to the expectations of central managers. For a decentralised organisation, an *optimal mechanism* is proposed whose instrument is the internal cost of capital. The same level of surplus is maintained by the conglomerate under both principles. When central managers do not know the types of subsidiaries, the estimations may be calibrated a posteriori — by looking for the functional types which are consistent with the preferred levels of equity. Thus, the mechanism promotes transparency within the institution. Moreover, the burden of arithmetic operations may be reduced if the distortion function is parametrically determined, such that a single real number accounts for the informational type[8].

Finally, the mechanism can be useful for regulatory purposes by determining the types which are consistent with the levels of risk capital observed in the industry. Institutions demanding the minimum capital requirement might be expecting a higher performance from their investments, than suggested by *average* knowledge. Moreover, though it is not possible to know when companies are underestimating their risk, as long as some information is private, rational decision makers reveal their type — for they maximise value.

[8] In Mierzejewski (2006) the model is presented in these terms, see also Wang (1995).

References

P. Albrecht. Risk based capital allocation. In J.L. Teugels and B. Sundt, editors, *Encyclopedia of Actuarial Science*, pages 1459–1466, New York, 2004. John Wiley & sons.

Basel Committee on Banking Supervision 1996. Amendment to the capital accord to incorporate market risks. http://www.bis.org, 1996.

Basel Committee on Banking Supervision 2004. International convergence of capital measurement and capital standards. a revised framework. http://www.bis.org, 2004.

J.D. Cummins and D.W. Sommer. Capital and risk property-liability insurance markets. *Journal of Banking and Finance*, 20:1069–1092, 1996.

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

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

D.W. Diamond and R.E. Verrecchia. Optimal managerial contracts and equilibrium security prices. *The Journal of Finance*, 37(2):275–287, 1982.

K.A. Froot, D.S. Scharfstein, and J.C. Stein. Risk management: coordinating corporate investment and financing policies. *The Journal of Finance*, 48(5):1629–1658, 1993.

M.J. Goovaerts, E. Van den Borre, and R.J.A. Laeven. Managing economic and virtual economic capital within financial conglomerates. *North American Actuarial Journal*, 9(3):77–89, 2005.

W.G. Hallerbach. Capital allocation, portfolio enhancement and performance measurement: a unified approach. http://www.gloriamundi.org, 2003.

M.C. Jensen. Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review*, 76(2):323–329, 1986.

R.J.A. Laeven and M.J. Goovaerts. An optimization approach to the dynamic allocation of economic capital. *Insurance: Mathematics and Economics*, 34(2):299–319, 2004.

R.C. Merton. An analytic derivation of the cost of deposit insurance and loan guarantees. *Journal of Banking and Finance*, 1:3–11, 1977.

R.C. Merton. A model of contract guarantees for credit-sensitive, opaque financial intermediaries. *Journal of Banking and Finance*, 1:1–13, 1997.

R.C. Merton and A.F. Perold. Theory of risk capital in financial firms. *Journal of Applied Corporate Finance*, 5:16–32, 1993.

F. Mierzejewski. Optimal capital allocation confronting bankruptcy and agency costs. *Bank- en Financiewezen*, pages 72–77, 2006.

M.H. Miller. The Modigliani-Miller propositions after thirty years. *Journal of Economic Perspectives*, 2(4):99–120, 1998.

F. Modigliani and M.H. Miller. The cost of capital, corporation finance and the theory of investment. *American Economic Review*, 49(4):655–69, 1959.

S.C. Myers and J.A. Read Jr. Capital allocation for insurance companies. *Journal of Risk and Insurance*, 68(4):545–580, 2001.

F. Saita. Risk capital aggregation: the risk managers perspective. http://www.gloriamundi.org, 2004.

J. Stiglitz. Some aspects of the pure theory of corporate finance: bankruptcies and take-overs. *Bell Journal of Economics and Management Science*, 3(2):458–482, 1972.

N.M. Stoughton and J. Zechner. Optimal capital allocation using RAROC and EVA. http://www.gloriamundi.org, 1999.

S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics & Economics*, 17:43–54, 1995.

S. Wang and V. Young. Ordering risks: Expected utility theory versus yaaris dual theory of risk. *Insurance: Mathematics & Economics*, 22:145–161, 1998.

S. Wang, V. Young, and H. Panjer. Axiomatic characterization of insurance prices. *Insurance: Mathematics & Economics*, 21:173–183, 1997.

ON THE RISK ADJUSTED PRICING METHODOLOGY MODEL FOR PRICING DERIVATIVE SECURITIES

Daniel Ševčovič

Department of Applied Mathematics and Statistics, Comenius University, Mlynská dolina, 842 48 Bratislava, Slovakia
Email: `sevcovic@fmph.uniba.sk`

Abstract

In this review paper we discuss a nonlinear model of Black-Scholes type for pricing derivative securities in the presence of both transaction costs as well as the risk from a volatile portfolio. The model is derived by following the Risk Adjusted Pricing Methodology approach proposed by Kratka (1998). It turns outs that prices of plain vanilla options can be computed from a solution to a fully nonlinear parabolic equation in which a diffusion coefficient representing volatility nonlinearly depends on the asset price and option's Gamma. It gives rise to explain several striking phenomena in option pricing analytically, including, in particular, the volatility smile behavior of the implied volatility.

1. INTRODUCTION

According to the classical theory due to Black, Scholes and Merton the price of an option in an idealized financial market can be computed from a solution to the well-known Black-Scholes linear parabolic equation (see e.g. Black and Scholes (1973), Kwok (1998), Dewynne et al. (1993), Hull (1989)). Assuming that the underlying asset follows a geometric Brownian motion one can derive a governing partial differential equation for the price of an option. We remind ourselves that the equation governing time evolution of the price $V(S,t)$ of an option is the following parabolic PDE:

$$\partial_t V + (r-q)S\partial_S V + \frac{1}{2}\hat{\sigma}^2 S^2 \partial_S^2 V - rV = 0 \tag{1}$$

where $\hat{\sigma}$ is a constant volatility of the underlying asset price process, $r > 0$ is the interest rate of a zero-coupon bond, $q \geq 0$ is the dividend yield rate. A solution $V = V(S,t)$ represents the price of an option at time $t \in [0,T]$ if the price of an underlying asset is $S > 0$. If the volatility $\hat{\sigma}$ is assumed to be constant the above equation is called the Black-Scholes equation derived by Black and Scholes (1973), and, independently by Merton (c.f. Kwok (1998)). The linear Black-Scholes equation has been derived under restrictive assumptions like e.g. perfect replication of a portfolio, frictionless, liquidity, complete markets, etc. Following this theory we can find a value of an option over moderate time intervals assuming transaction costs and the risk from a volatile portfolio are negligible. A solution to the linear Black-Scholes equation then provides a perfectly replicating hedging portfolio.

In recent years, some of these restrictive assumptions have been relaxed in order to model, for instance, the presence of transaction costs (Hoggard et al. (1994)), imperfect replication and investor's preferences (Barles and Soner (1998)), introduction of a given stock-trading strategy of a large trader (Frey and Patie (2002), Frey and Stremme (1997)), risk from unprotected portfolio (Kratka (1998), Jandačka and Ševčovič (2005)). These models lead to a generalized Black-Scholes equation for the price of an option in which the volatility need not be necessarily constant and it may depend on the asset price as well as the option price. More precisely, in these models the volatility has the general form:

$$\sigma^2 = \sigma^2(S^2\partial_S^2 V, S, T-t)\,. \tag{2}$$

For instance, if transaction costs are taken into account then the classical Black-Scholes theory is no longer applicable. In order to maintain the delta hedge one has to make frequent portfolio adjustments yielding thus a substantial increase in transaction costs. The effect of nontrivial transaction costs can be described by the so-called Leland model (cf. Hoggard et al. (1994)). In this model the volatility σ is given by $\sigma^2 = \hat\sigma^2(1 - \mathrm{Le}\,\mathrm{sgn}(\partial_S^2 V))$ where $\hat\sigma > 0$ is a constant historical volatility of the underlying asset price process and $\mathrm{Le} \geq 0$ is the so-called Leland constant given by $\mathrm{Le} = \sqrt{2/\pi} C/(\hat\sigma\sqrt{\Delta t})$. Here $C \geq 0$ is a constant round trip transaction cost per unit dollar of transaction in the assets market and $\Delta t > 0$ is the time-lag between portfolio adjustments. Since $S > 0$ we have

$$\sigma^2(S^2\partial_S^2 V, S, T-t) = \hat\sigma^2(1 - \mathrm{Le}\,\mathrm{sgn}(\partial_S^2 V))\,. \tag{3}$$

By assuming that investor's preferences are characterized by an exponential utility function, Barles and Soner (1998) derived a nonlinear Black-Scholes equation with the volatility σ given by

$$\sigma^2(S^2\partial_S^2 V, S, T-t) = \hat\sigma^2\left(1 + \Psi(a^2 e^{r(T-t)} S^2 \partial_S^2 V)\right)^2$$

where $a > 0$ is the risk-aversion coefficient and Ψ is a solution to the ODE: $\Psi'(x) = (\Psi(x) + 1)/(2\sqrt{x\Psi(x)} - x)$, $\Psi(0) = 0$. Another popular model has been derived for the case when the asset dynamics takes into account the presence of feedback effects. Frey and Stremme (1997) (see also Frey and Patie (2002)) introduced directly the asset price dynamics in the case when the large trader chooses a given stock-trading strategy. The volatility σ is nonconstant and it is given by:

$$\sigma^2(S^2\partial_S^2 V, S, T-t) = \hat\sigma^2\left(1 - \varrho S \partial_S^2 V\right)^{-2}$$

where $\hat\sigma, \varrho > 0$ are constants.

The last example of a nonlinear Black-Scholes equation is the so-called Risk Adjusted Pricing Methodology model proposed by Kratka (1998), revisited and modified by Jandačka and Ševčovič (2005). The idea of derivation of this model is simple: in order to maintain (imperfect) replication of a portfolio by the delta hedge one has to make frequent portfolio adjustments yielding thus a substantial increase in transaction costs. On the other hand, rare portfolio adjustments may lead to the increase of the risk from a volatile (unprotected) portfolio. Minimization of the sum of the measure of transaction costs and the risk from unprotected portfolio yields the optimal time lag between two consecutive portfolio adjustments. The resulting model is again a nonlinear Black-Scholes type equation with the volatility of the form

$$\sigma^2(S^2\partial_S^2 V, S, T-t) = \hat{\sigma}^2 \left(1 - \mu(S\partial_S^2 V)^{\frac{1}{3}}\right) \tag{4}$$

for $T - t > 0$ large enough where $\mu \geq 0$ is a coefficient proportional to the risk from volatile portfolio and transaction costs measures. In the next section we recall key steps and ideas of derivation of the Risk Adjusted Pricing Methodology (RAPM) model. We will furthermore present explanation of the volatility smile based on the RAPM model. We also discuss calibration of the RAPM model to real market data. We also introduce two new implied quantities: the implied RAPM volatility and implied RAPM risk coefficients. Finally, we will present results of calibration of these new implied quantities to real option and stock market data.

2. RISK ADJUSTED PRICING METHODOLOGY MODEL

In this section we recall key steps of derivation of the RAPM model. The original model was proposed by Kratka (1998). In Jandačka and Ševčovič (2005) we modified his approach (we chose a different measure for risk from unprotected portfolio) in order to construct a model which is scale invariant and mathematically well posed. These two important features were missing in the original model of Kratka. The model is based on the Black-Scholes parabolic PDE in which transaction costs are described by the Hoggard, Whalley and Wilmott extension of the Leland model (cf. Hoggard et al. (1994), Kwok (1998), Hull (1989)) whereas the risk from a volatile portfolio is described by the average value of the variance of the synthesized portfolio. Transaction costs as well as the volatile portfolio risk depend on the time-lag between two consecutive transactions. We define the total risk premium as a sum of transaction costs and the risk cost from the unprotected volatile portfolio. By minimizing the total risk premium functional we obtain the optimal length of the hedge interval. It also gives us a new strategy for hedging derivative securities based on option's Gamma parameter.

Concerning the dynamics of an underlying asset we will assume that the asset price $S = S(t), t \geq 0,$ follows a geometric Brownian motion with a drift ρ, standard deviation $\hat{\sigma} > 0$ and it may pay continuous dividends, i.e.

$$dS = (\rho - q)Sdt + \hat{\sigma}SdW \tag{5}$$

where dW denotes the differential of the standard Wiener process and $q \geq 0$ is a continuous dividend yield rate. This assumption is usually made when deriving the classical Black-Scholes equation (see e.g. Hull (1989), Kwok (1998)).

Similarly as in the derivation of the classical Black-Scholes equation we construct a synthesized portfolio Π consisting of a one option with a price V and δ assets with a price S per one asset:

$$\Pi = V + \delta S\,. \tag{6}$$

We recall that the key idea in the Black-Scholes theory is to examine the differential $\Delta\Pi$ of equation (6). The right-hand side of (6) can be differentiated by using Itô's formula whereas portfolio's increment $\Delta\Pi(t) = \Pi(t + \Delta t) - \Pi(t)$ of the left-hand side can be expressed as follows:

$$\Delta\Pi = r\Pi\Delta t + \delta q S \Delta t \tag{7}$$

where $r > 0$ is a risk-free interest rate of a zero-coupon bond. In the real world, such a simplified assumption is not satisfied and a new term measuring the total risk should be added to (7). More precisely, the change of the portfolio Π is composed of two parts: the risk-free interest rate part $r\Pi\Delta t$ and the total risk premium: $r_R S\Delta t$ where r_R is a risk premium per unit asset price. It means that $\Delta\Pi = r\Pi\Delta t + r_R S\Delta t$. The total risk premium r_R consists of the transaction risk premium r_{TC} and the portfolio volatility risk premium r_{VP}, i.e. $r_R = r_{TC} + r_{VP}$. Hence

$$\Delta\Pi = r\Pi\Delta t + \delta q S\Delta t + (r_{TC} + r_{VP}) S\Delta t\,. \tag{8}$$

Our next goal is to show how these risk premium measures r_{TC}, r_{VP} depend on the time lag and other quantities, like e.g. $\hat\sigma, S, V,$ and derivatives of V. The problem can be decomposed in two parts: modeling the transaction costs measure r_{TC} and volatile portfolio risk measure r_{VP}.

2.1. Modeling transaction costs and volatile portfolio risk measures

In practice, we have to adjust our portfolio by frequent buying and selling of assets. In the presence of nontrivial transaction costs, continuous portfolio adjustments may lead to infinite total transaction costs. A natural way how to consider transaction costs within the frame of the Black-Scholes theory is to follow the well known Leland approach extended by Hoggard, Whalley and Wilmott (cf. Hoggard et al. (1994), Kwok (1998)). In what follows, we recall crucial lines of the Hoggard, Whalley and Wilmott derivation of Leland's model in order to show how to incorporate the effect of transaction costs into the governing equation. More precisely, we will derive the coefficient of transaction costs r_{TC} occurring in (8).

Let us denote by C the round trip transaction cost per unit dollar of transaction. Then

$$C = (S_{ask} - S_{bid})/S \tag{9}$$

where S_{ask} and S_{bid} are the so-called Ask and Bid prices of the asset, i.e. the market price offers for selling and buying assets, respectively. Here $S = (S_{ask} + S_{bid})/2$ denotes the mid value.

In order to derive the term r_{TC} in (8) measuring transaction costs we will assume, for a moment, that there is no risk from the volatile portfolio, i.e. $r_{VP} = 0$. Then $\Delta V + \delta\Delta S = \Delta\Pi = r\Pi\Delta t + \delta q S\Delta t + r_{TC} S\Delta t$. Following Leland's approach (c.f. Hoggard et al. (1994)), using Itô's formula and assuming δ-hedging of a synthetised portfolio Π one can derive that the coefficient r_{TC} of transaction costs is given by the formula:

$$r_{TC} = \frac{C\hat\sigma S}{\sqrt{2\pi}} \left|\partial_S^2 V\right| \frac{1}{\sqrt{\Delta t}} \tag{10}$$

(see (Hoggard et al. 1994, Eq. (3)) and also formula (3)).

Next we focus our attention to the problem how to incorporate a risk from a volatile portfolio into the model. In the case when a portfolio consisting of options and assets is highly volatile an investor usually asks for a price compensation. Notice that exposure to risk is higher when the time-lag between portfolio adjustments is higher. We shall propose a measure of such a risk based on the volatility of a fluctuating portfolio. It can be measured by the variance of relative increments of the replicating portfolio $\Pi = V + \delta S$, i.e. by the term $var((\Delta\Pi)/S)$. Hence it is reasonable to define the measure r_{VP} of the portfolio volatility risk as follows:

$$r_{VP} = R\,\frac{var\left(\frac{\Delta\Pi}{S}\right)}{\Delta t}\,. \tag{11}$$

In other words, r_{VP} is proportional to the variance of a relative change of a portfolio per time interval Δt. A constant R is the so-called *risk premium coefficient*. It can be interpreted as the marginal value of investor's exposure to a risk. If we apply Itô's formula to the differential $\Delta\Pi = \Delta V + \delta\Delta S$ we obtain $\Delta\Pi = (\partial_S V + \delta)\,\hat\sigma S\Delta W + \frac{1}{2}\hat\sigma^2 S^2\Gamma(\Delta W)^2 + \mathcal{G}$ where $\Gamma = \partial_S^2 V$ and $\mathcal{G} = (\partial_S V + \delta)\rho S\Delta t + \partial_t V\Delta t$ is a deterministic term, i.e. $E(\mathcal{G}) = \mathcal{G}$ in the lowest order Δt-term approximation. Thus

$$\Delta\Pi - E(\Delta\Pi) = (\partial_S V + \delta)\,\hat\sigma S\phi\sqrt{\Delta t} + \frac{1}{2}\hat\sigma^2 S^2(\phi^2 - 1)\Gamma\Delta t$$

where ϕ is a random variable with the standard normal distribution such that $\Delta W = \phi\sqrt{\Delta t}$. Hence the variance of $\Delta\Pi$ can be computed as follows:

$$var(\Delta\Pi) = E\left([\Delta\Pi - E(\Delta\Pi)]^2\right) = E\left([(\partial_S V + \delta)\hat\sigma S\phi\sqrt{\Delta t} + \tfrac{1}{2}\hat\sigma^2 S^2\Gamma\left(\phi^2 - 1\right)\Delta t]^2\right)\,.$$

Similarly, as in the derivation of the transaction costs measure r_{TC} we assume δ-hedging of portfolio adjustments, i.e. we choose $\delta = -\partial_S V$. Since $E((\phi^2-1)^2) = 2$ we obtain an expression for the risk premium r_{VP} in the form:

$$r_{VP} = \frac{1}{2}R\hat\sigma^4 S^2\Gamma^2\Delta t\,. \tag{12}$$

Notice that in our approach the increase in the time-lag Δt between consecutive transactions leads to a linear increase of the risk from a volatile portfolio where the coefficient of proportionality depends on the asset price S, option's Gamma, $\Gamma = \partial_S^2 V$, as well as the constant historical volatility $\hat\sigma$ and the risk premium coefficient R.

2.2. Risk adjusted Black-Scholes equation

The total risk premium $r_R = r_{TC} + r_{VP}$ consists of two parts: transaction costs premium r_{TC} and the risk from a volatile portfolio r_{VP} premium defined as in (10) and (12), respectively. We assume that an investor is risk averse and he/she wants to minimize the value of the total risk premium r_R. For this purpose one has to choose the optimal time-lag Δt between two consecutive portfolio adjustments. As both r_{TC} as well as r_{VP} depend on the time-lag Δt so does the total risk premium r_R. In order to find the optimal value of Δt we have to minimize the following function:

$$\Delta t \mapsto r_R = r_{TC} + r_{VP} = \frac{C|\Gamma|\hat\sigma S}{\sqrt{2\pi}}\frac{1}{\sqrt{\Delta t}} + \frac{1}{2}R\hat\sigma^4 S^2\Gamma^2\Delta t\,.$$

The unique minimum of the function $\Delta t \mapsto r_R$ is attained at the time-lag $\Delta t_{opt} = K^2/(\hat{\sigma}^2|S\Gamma|^{\frac{2}{3}})$ where $K = (C/(R\sqrt{2\pi})^{\frac{1}{3}}$. For the minimal value of the function $\Delta t \mapsto r_R(\Delta t)$ we have

$$r_R(\Delta t_{opt}) = \frac{3}{2}\left(\frac{C^2 R}{2\pi}\right)^{\frac{1}{3}} \hat{\sigma}^2 |S\Gamma|^{\frac{4}{3}} . \tag{13}$$

Taking into account both transaction costs as well as risk effects from a volatile portfolio, we have shown that the equation for the change $\Delta\Pi$ of a portfolio Π reads as:

$$\Delta V + \delta \Delta S = \Delta \Pi \Delta t = r\Pi + \delta q S \Delta t + r_R S \Delta t$$

where r_R represents the total risk premium, $r_R = r_{TC} + r_{VP}$. On the other hand, by the no-arbitrage principle the change $\Delta\Pi$ in the portfolio Π is equal to the change $r\Pi\Delta t$ of secure bonds with the interest rate $r > 0$. Applying Itô's lemma to a smooth function $V = V(S,t)$ and assuming the δ-hedging strategy for the portfolio adjustments we finally obtain the following generalization of the Black-Scholes equation for valuing options:

$$\partial_t V + \frac{\hat{\sigma}^2}{2} S^2 \partial_S^2 V + (r-q) S \partial_S V - rV - r_R S = 0 .$$

By taking the optimal value of the total risk coefficient r_R derived as in (13), the option price V is a solution to the following nonlinear parabolic equation:

(*Risk adjusted Black-Scholes equation*)

$$\partial_t V + \frac{\hat{\sigma}^2}{2} S^2 \left(1 - \mu (S\partial_S^2 V)^{\frac{1}{3}}\right) \partial_S^2 V + (r-q) S \partial_S V - rV = 0 , \quad \text{where } \mu = 3\left(\frac{C^2 R}{2\pi}\right)^{\frac{1}{3}} . \tag{14}$$

In the case there are neither transaction costs ($C = 0$) nor the risk from a volatile portfolio ($R = 0$) we have $\mu = 0$. Then equation (14) reduces to the original Black-Scholes linear parabolic equation (1). We note that equation (14) is a backward parabolic PDE if and only if the function $\beta(H) = \frac{\hat{\sigma}^2}{2}(1 - \mu H^{\frac{1}{3}})H$ is an increasing function in the variable $H := S\Gamma = S\partial_S^2 V$. Hence, in order to verify parabolicity of (14), we have to assume the following condition:

$$S\partial_S^2 V(S,t) < \kappa := \left(\frac{3}{4\mu}\right)^3 . \tag{15}$$

If we consider prices of either Call or Put options computed from a solution to the classical Black-Scholes equation (1) then the term $S\Gamma = S\partial_S^2 V(S,t)$ becomes infinite at $S = E$ for $t \to T^-$ and the (15) condition is violated. The same feature is present in the generalized equation (14) yielding thus the change of the sign of the diffusion coefficient of (14) close to expiration time T. This is why we have to modify the model equation (14) near the expiration time, i.e. for $0 < T - t \ll 1$. The idea of modified early exercise behavior was introduced by Jandačka and Ševčovič (2005). It consists in determining the so-called switching time $t_* < T$ such that the RAPM model is modified as follows: the price of an option is given by a solution $V(S,t)$ to the following problem:

1. $V(S,t)$ is a solution to equation (14) on the time interval $0 < t < t_*$; whereas

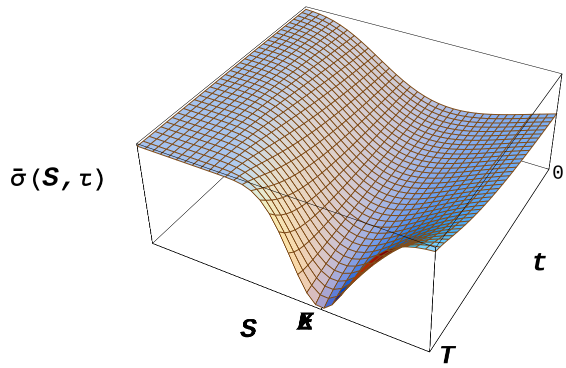


Figure 1: Explanation of the volatility smile based on RAPM. The implied volatility surface $(S,t) \mapsto \bar{\sigma}(S,t)$.

2. $V(S,t)$ is a solution to the linear Black-Scholes equation (1) on the time interval $t_* < t < T$ and satisfying the prescribed pay-off diagram at expiry $t = T$;

3. function $V(S,t)$ is continuous in $t = t_*$.

The switching time $t_* < T$ is chosen as nearest time to expiry T for which the value of $S\Gamma = S\partial_S^2 V$ is less or equal to the threshold value κ. Now if we compute the quantity $S\Gamma$ for plain Call or Put options by using the original Black-Scholes model (1) we obtain $\max_{S>0} S\Gamma(S,t_*) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2(T-t_*)}}$. Then we can deduce

$$T - t_* = \frac{C}{R\hat{\sigma}^2}\,. \tag{16}$$

As t_* must be positive we have $T - t_* < T$ it also turns out that we have to require the following structural condition

$$0 \le C < \hat{\sigma}^2 RT\,. \tag{17}$$

to be satisfied (see Jandačka and Ševčovič (2005) for details).

3. CALIBRATION OF THE RAPM MODEL TO REAL MARKET DATA

The purpose of this section is to discuss application of the RAPM model to real market option price data. We also introduce a concept of the so-called implied RAPM volatility σ_{RAPM} and the implied risk premium coefficient R. First we discuss capability of RAPM model to explain the so-called volatility smile analytically.

3.1. Volatility smile and RAPM model

One of the most striking phenomena in the Black-Scholes theory is the so-called *volatility smile* phenomenon. Notice that derivation of the classical Black-Scholes equation (1) relies on the assumption of a constant value of the volatility σ. On the other hand, as it might be documented by

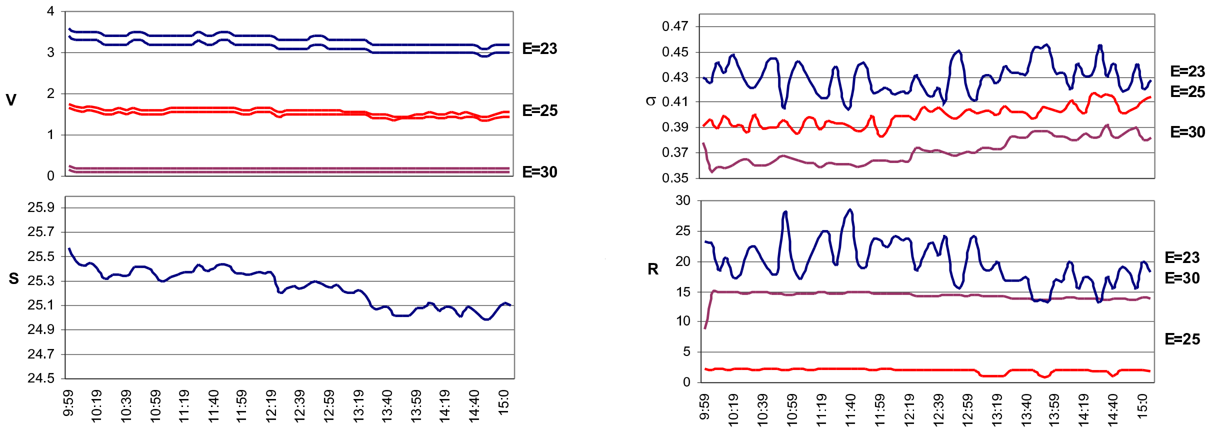


Figure 2: Intra-day behavior of Microsoft stocks (April 4, 2003) and shortly expiring Call options with expiry date April 19, 2003. Computed implied volatilities $\hat\sigma_{RAPM}$ and risk premium coefficients R.

many examples observed in market options data sets such an assumption is often violated. More precisely, the implied volatility σ_{impl} is no longer constant and it may depend on the asset price S, the strike price E as well as the time t.

In the RAPM approach we are able to explain the volatility smile analytically. The Risk adjusted Black-Scholes equation (14) can be viewed as an equation with a variable volatility coefficient, i.e. $\partial_t V + \frac{1}{2}\bar\sigma^2(S,t)\partial^2_S V + (r-q)S\partial_S V - rV = 0$ where $\Gamma = \partial^2_S V$ and the volatility $\bar\sigma^2(S,t)$ depends itself on a solution $V = V(S,t)$ as follows:

$$\bar\sigma^2(S,t) = \hat\sigma^2\left(1 - \mu(S\Gamma)^{1/3}\right) . \tag{18}$$

In Fig. 1 we show the dependence of the function $\bar\sigma(S,t)$ on the asset price S and time t. It should be obvious that the function $S \mapsto \bar\sigma(S,t)$ has a convex shape near the exercise price E. We have used the RAPM model in order to compute values of $\Gamma = \partial^2_S V$. We chose $\mu = 0.2, \hat\sigma = 0.3, r = 0.011$, and $T = 0.5$. In Fig. 1 we show the dependence of the function $\bar\sigma(S,t)$ on the asset price S and time t. It should be obvious that the function $S \mapsto \bar\sigma(S,t)$ has a convex shape near the exercise price E.

3.2. Implied volatility and risk premium in RAPM model

Let us denote $V(S,t;C,\hat\sigma,R)$ the value of a solution to (14) with parameters $C,\hat\sigma,R$. Suppose that the coefficient of transaction costs C is known from and is given by (9). In real option market data we can observe different Bid and Ask prices for an option, $V_{bid} < V_{ask}$, respectively. Let us denote by V_{mid} the mid value, i.e. $V_{mid} = \frac{1}{2}(V_{bid} + V_{ask})$. By the RAPM model we are able to explain such a Bid-Ask spread in option prices. The lower Bid price corresponds to a solution to the RAPM model with some nontrivial risk premium R whereas the mid value V_{mid} corresponds to a solution $V(S,t)$ for vanishing risk premium $R = 0$, i.e. to a solution of the linear Black-Scholes equation (1).

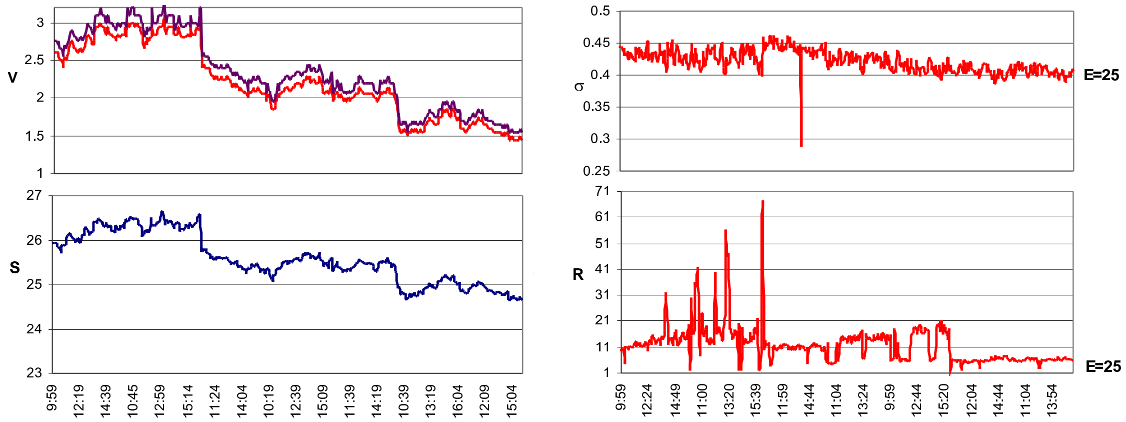


Figure 3: One week behavior of Microsoft stocks (March 20 - 27, 2003) and Call options with expiration date April 19, 2003. Computed implied volatilities $\hat\sigma_{RAPM}$ and risk premiums R.

In order to calibrate the RAPM model we are seeking for a couple $(\hat\sigma_{RAPM}, R)$ such that $V_{bid} = V(S,t;C,\hat\sigma_{RAPM},R)$ and $V_{mid} = V(S,t;C,\hat\sigma_{RAPM},0)$. It means that we have to find a solution to a nonlinear problem:

$$F(\hat\sigma, R) = (V_{bid}, V_{mid}) \tag{19}$$

where the mapping $F : \mathbb{R}^2 \to \mathbb{R}^2$ is defined as: $F(\hat\sigma,R) = (V(S,t;C,\hat\sigma,R), V(S,t;C,\hat\sigma,0))$. It can be solved numerically by means of the Newton-Kantorovich iterative method for solving algebraic equations. A solution $V(S,t;C,\hat\sigma,R)$ can be computed from the Risk adjusted Black-Scholes equation by means of finite difference (see Jandačka and Ševčovič (2005) for details).

As an example we considered sample data sets for Call options on Microsoft stocks. We considered a flat interest rate $r = 0.02$, a constant transaction cost coefficient $C = 0.01$ estimated from (9), and we assumed that the underlying asset pays no dividends, i.e. $q = 0$. In Fig. 2 we present results of calibration of implied couple $(\hat\sigma_{RAPM}, R)$. Interestingly enough, two Call options with higher strike prices $E = 25, 30$ had almost constant implied risk premium R. On the other the risk premium of an option with lowest $E = 23$ was fluctuating and it had highest average of R.

Finally, in Fig. 3 we present one week behavior of implied volatilities and risk premium coefficients for the Microsoft Call option on $E = 25$ expiring at $T =$ April 19, 2003. In the beginning of the investigated period the risk premium coefficient R was rather high and fluctuating. On the other hand, it tends to a flat value of $R \approx 5$ at the end of the week. Interesting feature can be observed at the end of the second day when both stock and option prices went suddenly down. The time series analysis of the implied volatility $\hat\sigma_{RAPM}$ from first two days was unable to predict such a behavior. On the other, high fluctuation in the implied risk premium R during first two days can send a signal to an investor that sudden changes can be expected in the near future.

4. CONCLUSIONS

In this paper we discussed the Risk Adjusted Pricing Methodology model for for pricing derivative securities in the presence of both transaction costs as well as the risk from unprotected portfolio. We showed that the option price can be deduced from a solution to a nonlinear parabolic PDE. The governing equation extends the classical Black-Scholes equation and Leland's equation to the case when the risk from unprotected portfolio is taken into account. We have performed extensive numerical testing of the model and compared the results to real option market data. Furthermore, we introduced a concept of the so-called implied RAPM volatility and implied risk premium coefficients. We have computed these implied quantities for sample option data sets and we have indicated how these implied factors can be used in qualitative analysis of option market data sets.

References

G. Barles and H.M. Soner. Option pricing with transaction costs and a nonlinear Black-Scholes equation. *Finance and Stochastics*, 2(4):369–397, 1998.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

J.N. Dewynne, S.D. Howison, J. Rupf, and P. Wilmott. Some mathematical results in the pricing of american options. *European Journal of Applied Mathematics*, 4:381–398, 1993.

R. Frey and P. Patie. Risk management for derivatives in illiquid markets: A simulation study. In K. Sandmann and Ph. Schönbucher, editors, *Advances in Finance and Stochastics*, pages 137–159. Springer, Berlin, 2002.

R. Frey and A. Stremme. Market volatility and feedback effects from dynamic hedging. *Mathematical Finance*, 4:351–374, 1997.

T. Hoggard, A. E. Whalley, and P. Wilmott. Hedging option portfolios in the presence of transaction costs. *Advances in Futures and Options Research*, 7:21–35, 1994.

J. Hull. *Options, Futures and Other Derivative Securities*. Prentice Hall, New-York, 1989.

M. Jandačka and D. Ševčovič. On the risk adjusted pricing methodology based valuation of vanilla options and explanation of the volatility smile. *Journal of Applied Mathematics*, 3:235–258, 2005.

M. Kratka. No mystery behind the smile. *Risk*, 9:67–71, 1998.

Y. K. Kwok. *Mathematical Models of Financial Derivatives*. Springer, New York, Heidelberg, Berlin, 1998.

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 "4th Actuarial and Financial Mathematics Day" (10 februari 2006, Prof. M. Vanmaele)

De "4th Actuarial and Financial Mathematics Day" is een vaste waarde geworden als contactforum. Niet alleen academici maar ook heel wat collega's uit de bank- en verzekeringswereld blijven de weg vinden naar dit jaarlijkse evenement. Het is de gelegenheid bij uitstek om op de hoogte te blijven van het recente onderzoek op het vlak van financiële en actuariële wiskunde in België en van nieuwe uitdagingen die ons te wachten staan zoals in het kader van Basel II. Naast twee gastsprekers kwamen doctoraatsstudenten, postdocs en mensen uit de bedrijfswereld aan bod. In deze publicatie vindt u een neerslag van de voorgestelde onderwerpen. Alle onderwerpen kunnen gesitueerd worden in het ruime gebied van financiële en actuariële toepassingen van wiskunde, maar met een grote variatie: de bijdragen betreffen "capital allocation" problemen, modellen voor kredietrisico, voor stop-loss premies en voor basket- en spreadopties, risicomanagement van coupon bonds, etc.