KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

5TH ACTUARIAL AND FINANCIAL MATHEMATICS DAY

February 9, 2007

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

CONTACTFORUM

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

5TH ACTUARIAL AND FINANCIAL MATHEMATICS DAY

February 9, 2007

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

CONTACTFORUM

Handelingen van het contactforum "5th Actuarial and Financial Mathematics Day" (9 februari 2007, 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/2007/0455/00

Printed by *Universa Press, 9230 Wetteren, Belgium*

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

5th Actuarial and Financial Mathematics Day

PREFACE

The fifth edition of the Contactforum “Actuarial and Financial Mathematics Day” attracted many participants, both researchers and practitioners. We welcomed this year several participants from abroad, indicating that this event is becoming to be known internationally.
This contactforum aims to bring together young researchers, in particular Ph.D. students and Postdocs, working in the field of Financial and Actuarial Mathematics to discuss recent developments in the theory of mathematical finance and insurance and its application to current issues faced by the industry and to identify the substantive problems confronting academic researchers and finance professionals. We provide a forum for the discussion of advances within this field. In particular, we want to promote the exchange of ideas between practitioners and academics.
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 risk measures is included in these transactions.

We thank all our speakers and discussants (*Jasper Anderluh, Katrien Antonio, Griselda Deelstra, Henrik Jönsson, Nele Vandaele, Maarten Van Wieren, David Vyncke)*, for their enthusiasm and their interesting contributions which made this day a great success. 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.

We plan a two day international event for the next meeting in 2008 with the focus on the interplay between finance and insurance.

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

KONINKLIJKE VLAAMSE ACADEMIE VAN BELGIE
VOOR WETENSCHAPPEN EN KUNSTEN

5th Actuarial and Financial Mathematics Day

CONTENTS

Invited talk

Contributed talks

INVITED TALK

CAPITAL ALLOCATION WITH RISK MEASURES

Andreas Tsanakas

Faculty of Actuarial Science and Insurance, Cass Business School, City University, 106 Bunhill Row, London EC1Y 8TZ, UK.
Email: a.tsanakas.1@city.ac.uk

Abstract

This brief review paper covers the use of risk measures for assessing economic capital requirements and considers the problem of allocating aggregate capital to sub-portfolios.

1. INTRODUCTION

A *risk measure* is a function that assigns real numbers to random variables representing uncertain pay-offs, e.g. insurance losses. The interpretation of a risk measure's outcome depends on the context in which it is used. Historically there have been three main areas of application of risk measures:

- As representations of risk aversion in asset pricing models, with a leading paradigm the use of the variance as a risk measure in Markowitz portfolio theory, see Markowitz (1952).
- As tools for the calculation of the insurance price corresponding to a risk. Under this interpretation, risk measures are called *premium calculation principles* in the classic actuarial literature, e.g. Goovaerts et al. (1984).
- As quantifiers of the economic capital that the holder of a particular portfolio or risks should safely invest in e.g. Artzner et al. (1999).

This contribution is mainly concerned with the latter interpretation of risk measures.

The *economic* or *risk capital* held by a (re)insurer corresponds to the level of safely invested assets used to protect itself against unexpected volatility of its portfolio's outcome. One has to distinguish economic capital from regulatory capital, which is the minimum required economic capital level as set by the regulator. In fact, much of the impetus for the use of risk measures in the quantification of capital requirements comes from the area of regulating financial institutions.

Banking supervision (Basel Committee on Banking Supervision) and, increasingly, insurance regulation (European Commission) have been promoting the development of companies' *internal models* for modelling risk exposures. In that context, the application of a risk measure (most prominently *Value-at-Risk*) on the modelled aggregate risk profile of the insurance company is required.

Economic capital generally exceeds the minimum set by the regulator. Subject to that constraint, economic capital is determined so as to maximise performance metrics for the insurance company, such as total shareholder return (Exley and Smith 2006). Such maximisation takes into account two conflicting effects of economic capital (Hancock et al. 2001):

- An insurance company's holding economic capital incurs costs for its shareholders, which can be opportunity or frictional costs.
- Economic capital reduces the probability of default of the company as well as the severity of such default on its policyholders. This enables the insurance company to obtain a better rating of its financial strength and thereby attract more insurance business at higher prices.

Calculation of the optimal level of economic capital using such arguments is quite complicated and depends on factors that are not always easy to quantify, such as frictional capital costs, and on further constraints, such as the ability of an insurance company to raise capital in a particular economic and regulatory environment.

We could however consider that there is a particular calibration of the (regulatory or other) risk measure, which gives for the insurance company's exposure a level of economic capital that coincides with that actually held by the company. In that sense risk measures can be used to interpret exogenously given economic capital amounts. Such interpretation can be in the context of capital being set to achieve a target rating, often associated with a particular probability of default. Discussion of economic capital in the context of risk measures should therefore be caveated as being *ex-post*.

Finally, we note that the level of economic capital calculated by a risk measure may be a notional amount, as the company will generally not invest all its surplus in risk-free assets. This can be dealt with by absorbing the volatility of asset returns in the risk capital calculation itself.

2. DEFINITION AND EXAMPLES OF RISK MEASURES

We consider a set of risks $\mathcal{X}$ that the insurance company can be exposed to. The elements $X \in \mathcal{X}$ are random variables, representing losses at a fixed time horizon T. If under a particular state of the world ω the variable $X(\omega) > 0$ we will consider this to be a loss, while negative outcomes will be considered as gains. For convenience it is assumed throughout that the return from risk-free investment is 1 or alternatively that all losses in $\mathcal{X}$ are discounted at the risk-free rate. A risk measure ρ is then defined as a functional

$$\rho : \mathcal{X} \mapsto \mathbb{R}. \tag{1}$$

If X corresponds to the aggregate net risk exposure of an insurance company (i.e. the difference between liabilities and assets, excluding economic capital) and economic capital corresponds to $\rho(X)$, then we assume that the company defaults when $X > \rho(X)$.

In the terminology of Artzner et al. (1999) (and subject to some simplification), a risky position X is called acceptable if $\rho(X) < 0$, implying that some capital may be released without endangering the security of the holder of X, while $\rho(X) \geq 0$ means that X is a non-acceptable position and that some capital has to be added to it.

Some examples of simple risk measures proposed in the actuarial and financial literature (e.g. Bühlmann (1970), Denuit et al. (2005)) are as follows.

Example 1 (Expected value principle)

$$\rho(X) = \lambda E[X],\ \lambda \geq 1. \tag{2}$$

Besides its application in insurance pricing, where it represents a proportional loading, this risk measure in essence underlies simple regulatory minimum requirements, such as the current EU Solvency rules, which determine capital as a proportion of an exposure measure such as premium.

Example 2 (Standard deviation principle)

$$\rho(X) = E[X] + \kappa\sigma[X],\ \kappa \geq 0. \tag{3}$$

In this case the loading is risk-sensitive, as it is a proportion of the standard deviation. This risk measure is encountered in reinsurance pricing, while also relating to Markowitz portfolio theory. In the context of economic capital, it is usually derived as an approximation to other risk measures, with this approximation being accurate for the special case of multivariate normal (more generally elliptical) distributions (Embrechts et al. 2002).

Example 3 (Exponential Premium Principle)

$$\rho(X) = \frac{1}{a} \ln E[e^{aX}],\ a > 0. \tag{4}$$

The exponential premium principle is a very popular risk measure in the actuarial literature, e.g. Gerber (1974). Part of the popularity stems from the fact that, in the classic ruin problem, it gives the required level of premium associated with Kramer-Lundberg bounds for ruin probabilities. We note that this risk measure has been recently considered in the finance literature under the name ‘entropic risk measure’ (Föllmer and Schied 2002b).

Example 4 (Value-at-Risk)

$$\rho(X) = \mathit{VaR}_p(X) = F_X^{-1}(p),\ p \in (0,1), \tag{5}$$

where F_X is the cumulative probability distribution of X and F_X^{-1} is its (pseudo-)inverse. $\mathrm{VaR}_p(X)$ is easily interpreted as the amount of capital that, when added to the risk X, limits the probability of default to $1 - p$. Partly because of its intuitive attractiveness Value-at-Risk has become the risk measure of choice for both banking and insurance regulators. For example, the UK regulatory regime for insurers uses $\mathrm{VaR}_{0.995}(X)$ (Financial Services Authority), while a similar risk measure has been be proposed in the context of the new EU-wide Solvency II regime (European Commission).

Example 5 (Expected Shortfall)

$$\rho(X) = ES_p(X) = \int_p^1 F_X^{-1}(q)dq,\ p \in (0,1). \tag{6}$$

This risk measure, also known as Tail-(or Conditional-)Value-at-Risk, corresponds to the average of all VaR_ps above the threshold p. Hence it reflects both the probability and the severity of a potential default. Expected shortfall has been proposed in the literature as a risk measure correcting some of the theoretical weaknesses of Value-at-Risk (Wirch and Hardy 1999). Subject to continuity of F_X at the threshold VaR_p, Expected Shortfall coincides with the *Tail Conditional Expectation*, defined by

$$\rho(X) = E[X|X > F_X^{-1}(p)]. \tag{7}$$

Example 6 (Distortion risk measure)

$$\rho(X) = -\int_{-\infty}^0 (1 - g(1 - F_X(x)))dx + \int_0^\infty g(1 - F_X(x))dx, \tag{8}$$

where $g : [0,1] \mapsto [0,1]$ is increasing and concave (Wang 1996). This risk measure can be viewed as an expectation under a distortion of the probability distribution effected by the function g. It can be easily shown that Expected Shortfall is a special case obtained by a bilinear distortion (Wirch and Hardy 1999). Distortion risk measures can be viewed as Choquet integrals (Denneberg (1990), Denneberg (1994)), which are extensively used in the economics of uncertainty, e.g. Schmeidler (2003). An equivalent class of risk measures defined in the finance literature are known as *spectral risk measures* (Acerbi 2002).

Example 7 (Distortion-exponential risk measure)

$$\rho(X) = \frac{1}{a}\ln[\rho_*(e^{aX})], \tag{9}$$

where ρ_* is a distortion risk measure. This risk measure was proposed in Tsanakas and Desli (2003) and it combines the properties of the exponential premium principle with those of distortion risk measures.

3. PROPERTIES OF RISK MEASURES

The literature is rich in discussions of the properties of alternative risk measures, as well as the desirability of such properties, e.g. Goovaerts et al. (1984), Artzner et al. (1999), Goovaerts et al. (2003), Denuit et al. (2005). In view of this, the current discussion is invariably selective.

An often required property of risk measures is that of *monotonicity*, stating

$$\text{If } X \le Y, \text{ then } \rho(X) \le \rho(Y). \tag{10}$$

This reflects the obvious requirement that losses that are always higher should also attract a higher capital requirement.

A further appealing property is that of *translation* or *cash invariance*,

$$\rho(X + a) = \rho(X) + a, \text{ for } a \in \mathbb{R}. \tag{11}$$

This postulates that adding a constant loss amount to a portfolio increases the required risk capital by the same amount. We note that this has the implication that

$$\rho(X - \rho(X)) = \rho(X) - \rho(X) = 0, \tag{12}$$

which, in conjunction with monotonicity, facilitates the interpretation of $\rho(X)$ as the minimum capital amount that has to be added to X in order to make it acceptable.

Two conceptually linked properties are the ones of *positive homogeneity*,

$$\rho(bX) = b\rho(X), \text{ for } b \geq 0, \tag{13}$$

and *subadditivity*,

$$\rho(X + Y) \leq \rho(X) + \rho(Y), \text{ for all } X, Y \in \mathcal{X}. \tag{14}$$

Positive homogeneity postulates that a linear increase in the risk exposure X also implies linear increase in risk. Subadditivity requires that the merging of risks should always yield a reduction in rick capital due to diversification.

Risk measures satisfying the four properties of monotonicity, translation invariance, positive homogeneity and subadditivity have become widely known as *coherent* (Artzner et al. 1999). This particular axiomatization, also proposed in an actuarial context (Denneberg (1990), Wang et al. (1997)), has achieved near-canonical status in the world of financial risk management. While Value-at-Risk generally fails the subadditivity property, due to its disregard for the extreme tails of distributions, part of its appeal to regulators and practitioners stems of its use as an approximation to a coherent risk measure.

Nonetheless, coherent risk measures have also attracted criticism because of their insensitivity to the aggregation of large positively dependent risks implied by the latter two properties, e.g. Goovaerts et al. (2003). The weaker property of *convexity* has been proposed in the literature (Föllmer and Schied 2002a), a property already discussed in Deprez and Gerber (1985). Convexity requires that:

$$\rho(\lambda X + (1 - \lambda)Y) \leq \lambda\rho(X) + (1 - \lambda)\rho(Y), \text{ for all } X, Y \in \mathcal{X} \text{ and } \lambda \in [0, 1]. \tag{15}$$

Convexity, while retaining the diversification property, relaxes the requirement that a risk measure must be insensitive to aggregation of large risks. It is noted that subadditivity is obtained by combining convexity with positive homogeneity. Risk measures satisfying convexity and applying increasing penalties for large risks have been proposed in Tsanakas and Desli (2003).

Risk measures produce an ordering of risks, in the sense that $\rho(X) \leq \rho(Y)$ means that X is considered less risky than Y. One would wish that ordering to conform to standard economic theory, i.e. to be consistent with widely accepted notions of stochastic order such as 1st and 2nd order stochastic dominance and convex order, see Müller and Stoyan (2002), Denuit et al. (2005). It has been shown that under some relatively mild technical conditions, risk measures that are monotonic and convex produce such a consistent ordering of risks (Bäuerle and Müller 2006).

A further key property relates to the dependence structure between risks under which the risk measure becomes *additive*

$$\rho(X+Y)=\rho(X)+\rho(Y), \tag{16}$$

as this implies a situation where neither diversification credits nor aggregation penalties are assigned. In the context of subadditive risk measures, *comonotonic additivity* is a sensible requirement, as it postulates that no diversification is applied in the case of comonotonicity (the maximal level of dependence between risks, e.g. Dhaene et al. (2002)). On the other hand, one could require a risk measure to be *independent additive*. If such a risk measure is also consistent with the stop-loss or convex order, by the results of Dhaene and Goovaerts (1996), it is guaranteed to penalize any positive dependence by being superadditive (i.e. $\rho(X+Y)\geq\rho(X)+\rho(Y)$) and reward any negative dependence by being subadditive.

The risk measures defined above satisfy the following properties:

Expected value principle Monotonic, positive homogenous, additive for all dependence structures.

Standard deviation principle Translation invariant, positive homogenous, subadditive.

Exponential premium principle Monotonic, translation invariant, convex, independent additive.

Value-at-Risk Monotonic, translation invariant, positive homogenous, subadditive for joint-elliptically distributed risks (Embrechts et al. 2002), comonotonic additive.

Expected Shortfall Monotonic, translation invariant, positive homogenous, subadditive, comonotonic additive.

Distortion risk measure Monotonic, translation invariant, positive homogenous, subadditive, comonotonic additive.

Distortion-exponential risk measure Monotonic, translation invariant, convex.

Finally we note that all risk measures discussed in this contribution are *law invariant*, meaning that $\rho(X)$ only depends on the distribution function of X (Wang et al. (1997), Kusuoka (2001)). This implies that two risks characterised by the same probability distribution would be allocated the same amount of economic capital.

4. CONSTRUCTIONS AND REPRESENTATIONS OF RISK MEASURES

4.1. Indifference arguments

Economic theories of choice under risk seek to model the preferences of economic agents with respect to uncertain pay-offs. They generally have representations in terms of *preference functionals* $V:-\mathcal{X}\mapsto\mathbb{R}$, in the sense that

$$-X \text{ is preferred to } -Y \quad\Leftrightarrow\quad V(-X)\geq V(-Y). \tag{17}$$

(Note that the minus sign is applied because we have defined risk as losses, while preference functionals are typically applied on pay-offs.)

Then a risk measure can be defined by assuming that the addition to initial wealth W of a liability X and the corresponding capital amount $\rho(X)$ does not affect preferences (Bühlmann 1970)

$$V(W_0 - X + \rho(X)) = V(W_0). \tag{18}$$

Often in this context $W_0 = 0$ is assumed for simplicity.

The leading paradigm of choice under risk is the von Neumann-Morgenstern *expected utility theory* (Von Neumann and Morgenstern 1944), under which

$$V(W) = E[u(W)], \tag{19}$$

where u is an increasing and concave *utility function*. A popular choice of utility function is the *exponential utility*

$$u(w) = \frac{1}{a}\left(1 - e^{-aw}\right),\ a > 0. \tag{20}$$

It can be easily seen that equations (18), (19) and (20) yield the exponential premium principle defined in section 2.

An alternative theory is the *dual theory of choice under risk* (Yaari 1987), under which

$$V(W) = -\int_{-\infty}^{0}(1 - h(1 - F_W(w)))dw + \int_{0}^{\infty} h(1 - F_W(w))dw, \tag{21}$$

where $h : [0,1] \mapsto [0,1]$ is increasing and convex. It can then be shown that the risk measure obtained from (18) and (21) is a distortion risk measure with $g(s) = 1 - h(1-s)$. For the function

$$h(s) = 1 - (1-s)^{\frac{1}{\gamma}},\ \gamma > 1 \tag{22}$$

the well known *proportional hazards transform* with $g(s) = s^{\frac{1}{\gamma}}$ is obtained (Wang 1996).

More detailed discussions of risk measures resulting from alternative theories of choice under risk and references to the associated economics literature are given in Tsanakas and Desli (2003), Denuit et al. (2006).

It should also be noted that the construction of risk measures from economic theories of choice must not necessarily be via indifference arguments. If a risk measure satisfies the convexity and monotonicity properties, then by setting $U(W) = -\rho(-W)$ we obtain a monotonic concave preference functional. The translation invariance property of the risk measure then makes U also translation invariant. Hence we could consider convex risk measures as the subset of concave preference functionals that satisfy the translation invariance property (subject to a minus sign). Such preference functionals are sometimes called *monetary utility functions*, as their output can be interpreted as being in units of money rather than of an abstract notion of satisfaction.

4.2. Axiomatic characterisations

An alternative approach to deriving risk measures is by fixing a set of properties that risk measures should satisfy and then seeking an explicit functional representation.

For example, coherent (i.e. monotonic, translation invariant, positive homogenous and subadditive) risk measures can be represented by (Artzner et al. 1999)

$$\rho(X) = \sup_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}}[X], \tag{23}$$

where $\mathcal{P}$ is a set of probability measures. By adding the comonotonic additivity property one gets the more specific structure of $\mathcal{P} = \{\mathbb{P} : \mathbb{P}(A) \leq v(A) \text{ for all sets } A\}$, where v is a submodular set function known as a (Choquet) capacity (Denneberg 1994). The additional property of law invariance enables writing $v(A) = g(\mathbb{P}_0(A))$ where $\mathbb{P}_0$ is the objective probability measure and g a concave distortion function (Wang et al. 1997). This finally yields a representation of coherent, comonotonic additive, law invariant risk measures as distortion risk measures. An alternative route towards this representation is given by Kusuoka (2001).

The probability measures in $\mathcal{P}$ have been termed *generalized scenarios* (Artzner et al. 1999) with respect to which the worst case expected loss is considered. On the other hand, representations such as (23) have been derived in the context of robust statistics (Huber 1981) and decision theory, known as the multiple-priors model (Gilboa and Schmeidler 1989).

A related representation result for convex risk measures is derived in Föllmer and Schied (2002a), while results for independent additive risk measures are given in Gerber and Goovaerts (1981), Goovaerts et al. (2004).

4.3. Re-weighting probabilities

An intuitive construction of risk measures is by re-weighting the probability distribution of the underlying risk

$$\rho(X) = E[X\zeta(X)], \tag{24}$$

where ζ is generally assumed to be an increasing function with $E[\zeta(X)] = 1$ and representation (24) could be viewed as an expectation under a change of measure. Representation (24) is particularly convenient when risk measures and related functionals have to be evaluated by Monte-Carlo simulation.

Many well-known risk measures can be obtained in this way. For example, making appropriate assumptions on F_X and g one can easily show that for distortion measures it is

$$\rho(X) = E[Xg'(1 - F_X(X))]. \tag{25}$$

On the other hand the exponential principle can be written as:

$$\rho(X) = E\left[X \int_0^1 \frac{e^{\gamma a X}}{E[e^{\gamma a X}]} d\gamma\right]. \tag{26}$$

The latter representation is sometimes called a 'mixture of Esscher principles' and studied in more generality in Gerber and Goovaerts (1981), Goovaerts et al. (2004).

5. CAPITAL ALLOCATION

5.1. Problem definition

Often the requirement arises that the risk capital calculated for an insurance portfolio has to be allocated to business units. There may be several reasons for such a capital allocation exercise, the main ones being performance measurement / management and insurance pricing.

Capital allocation is not a trivial exercise, given that in general the risk measure used to set the aggregate capital is not additive. In other words, if one has an aggregate risk Z for the insurance company, breaking down to sub-portfolios $X_1, \ldots, X_n$, such that

$$Z = \sum_{j=1}^{n} X_j, \tag{27}$$

it generally is

$$\rho(Z) \neq \sum_{j=1}^{n} \rho(X_j), \tag{28}$$

due to diversification / aggregation issues.

The capital allocation problem then consists of finding constants $d_1, \ldots, d_n$ such that

$$\sum_{j=1}^{n} d_j = \rho(Z), \tag{29}$$

where the allocated capital amount d_i should in some way reflect the risk of sub-portfolio X_i. Early papers in the actuarial literature that deal with cost allocation problems in insurance are Bühlmann (1996) and Lemaire (1984), the former taking a risk theoretical view, while the later examining alternative allocation methods from the perspective of cooperative game theory. A specific application of cooperative game theory to risk capital allocation, including a survey of the relevant literature, is Denault (2001).

5.2. Marginal cost approaches

Marginal cost approaches associate allocated capital to the impact that changes in the exposure to sub-portfolios have on the aggregate capital. Denote for vector of weights $\mathbf{w} \in [0,1]^n$,

$$Z^{\mathbf{w}} = \sum_{j=1}^{n} w_j X_j. \tag{30}$$

Then the marginal cost of each sub-portfolio is given by

$$\mathrm{MC}(X_i; Z) = \left. \frac{\partial \rho(Z^{\mathbf{w}})}{\partial w_i} \right|_{\mathbf{w}=\mathbf{1}}, \tag{31}$$

subject to appropriate differentiability assumptions. If the risk measure is positive homogeneous, then by Euler's theorem we have that

$$\sum_{j=1}^{n} \mathrm{MC}(X_j; Z) = \rho(Z) \tag{32}$$

and we can hence use marginal costs directly $d_i = \mathrm{MC}(X_i; Z)$ as the capital allocation.

If the risk measure is in addition subadditive then we have that (Aubin 1981):

$$d_i = \mathrm{MC}(X_i; Z) \leq \rho(X_i), \tag{33}$$

i.e. the allocated capital amount is always lower than the stand-alone risk capital of the sub-portfolio. This corresponds to the game theoretical concept of the *core*, in that the allocation does not provide an incentive for splitting the aggregate portfolio. This requirement is consistent with the subadditivity property, which postulates that there is always a benefit in pooling risks.

In the case that no such strong assumptions as positive homogeneity (and subadditivity) are made with respect to the risk measure, marginal costs will in general not yield an appropriate allocation, as they will not add up to the aggregate risk. Cooperative game theory then provides an alternative allocation method, based on the *Aumann-Shapley value* (Aumann and Shapley 1974), which can be viewed as a generalisation of marginal costs

$$\mathrm{AC}(X_i, Z) = \int_0^1 \mathrm{MC}(X_i; \gamma Z) d\gamma. \tag{34}$$

It can easily be seen that if we set $d_i = \mathrm{AC}(X_i, Z)$ then the d_is add up to $\rho(Z)$ and that for positive homogenous risk measures the Aumann-Shapley allocation reduces to marginal costs. Early applications of the Aumann-Shapley value to cost allocation problems are Billera and Heath (1982), Mirman and Tauman (2006).

For the examples of risk measures that were introduced in section 2, the following allocations are obtained from marginal costs / Aumann-Shapley.

Example 8 (Expected value principle)

$$d_i = \lambda E[X_i] \tag{35}$$

Example 9 (Standard deviation principle)

$$d_i = E[X_i] + \kappa \frac{\mathit{Cov}(X_i, Z)}{\sigma[Z]} \tag{36}$$

Example 10 (Exponential Premium Principle)

$$d_i = \int_0^1 \frac{E[X_i \exp(\gamma a Z)]}{E[\exp(\gamma a Z)]} d\gamma \tag{37}$$

Example 11 (Value-at-Risk (Tasche 2004))

$$d_i = E[X_i | Z = \mathit{VaR}_p(Z)] \tag{38}$$

under suitable assumptions on the joint probability distribution of (X_i, Z).

Example 12 (Expected Shortfall (Tasche 2004))

$$d_i = E[X_i | Z > \textit{VaR}_p(Z)], \tag{39}$$

under suitable assumptions on the joint probability distribution of (X_i, Z).

Example 13 (Distortion risk measure (Tsanakas 2004))

$$d_i = E[X_i g'(1 - F_Z(Z))] \tag{40}$$

assuming representation (25) is valid.

Example 14 (Distortion-exponential Risk Measures)

$$d_i = \int_0^1 \frac{E[X_i \exp(\gamma a Z) g'(1 - F_Z(Z))]}{E[\exp(\gamma a Z) g'(1 - F_Z(Z))]} d\gamma \tag{41}$$

5.3. Alternative approaches

While marginal cost-based approaches are well-established in the literature, there are a number of alternative approaches to capital allocation. For example, we note that marginal costs generally depend on the joint distribution of the individual sub-portfolio and the aggregate risk. In some cases this dependence may not be desirable, for example when one tries to measure the performance of sub-portfolios to allocate bonuses. In that case, a simple *proportional repartition of costs* (Lemaire 1984) may be appropriate:

$$d_i = \rho(X_i) \frac{\rho(Z)}{\sum_{j=1}^n \rho(X_j)}. \tag{42}$$

Different issues emerge when the capital allocation is to be used for managing the performance of the aggregate portfolio, as measured by a particular metric such as return-on-capital. Assume that $\hat{X}_i$, $i = 1, \ldots, n$ correspond to the liabilities from sub-portfolio i minus reserves corresponding to those liabilities, such that $E[\hat{X}_i] = 0$. We then have the breakdown

$$X_i = \hat{X}_i - p_i, \tag{43}$$

where p_i corresponds to the underwriting profit from the insurer's sub-portfolio (e.g. line of business) i, such that $\sum_{j=1}^n \hat{X}_j = \hat{Z}$ and $\sum_{j=1}^n p_j = p$. Then we define the return on capital for the whole insurance portfolio by

$$\text{RoC} = \frac{p}{\rho(\hat{Z})}. \tag{44}$$

This is discussed in depth in Tasche (2004) for the case that ρ is a coherent risk measure. It is then considered whether assessing the performance of sub-portfolios by

$$\text{RoC}_i = \frac{p_i}{d_i}, \tag{45}$$

where d_i represents capital allocated to $\hat{X}_i$, provides the right incentives for optimizing performance. It is shown that marginal costs is the unique allocation mechanism that satisfies this requirement as set out in that paper. A closely related argument is that under the marginal cost allocation a portfolio balanced to optimize aggregate return on capital has the property that $\mathrm{RoC} = \mathrm{RoC}_i$ for all i. While this produces a useful performance yardstick that can be used throughout the company, some care has to be taken when applying marginal cost methodologies. In particular, if the marginal capital allocation to a sub-portfolio is small e.g. for reasons of diversification, the insurer should be careful not to let that fact undermine underwriting standards. A proportional allocation method could also be used for reference, to avoid that danger.

Often one may be interested in calculating capital allocations that are in some sense optimal. For example, in Dhaene et al. (2005) capital allocations are calculated such that a suitably defined distance function between individual sub-portfolios and allocated capital levels is minimized. This methodology reproduces many capital allocation methods found in the literature, while also considering the case that aggregate economic capital is exogenously given rather than calculated via a risk measure. A different optimization approach to capital allocation is presented in Laeven and Goovaerts (2003).

An alternative strand of the literature on capital allocation relates to the pricing of the policyholder deficit (also known as the 'limited liability put option'), due to the insurer's potential default (Myers and Read 2001) and considers the marginal impact of subportfolios on the market price of the deficit.

References

C. Acerbi. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26(7):1505–1518, 2002.

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

J-P. Aubin. Cooperative fuzzy games. *Mathematics of Operations Research*, 6(1):1–13, 1981.

R.J. Aumann and L.S. Shapley. *Values of Non-Atomic Games*. Princeton University Press, Princeton, 1974.

Basel Committee on Banking Supervision. The New Basel Capital Accord, 2003.

N. Bäuerle and A. Müller. Stochastic orders and risk measures: Consistency and bounds. *Insurance: Mathematics and Economics*, 38(1):132–148, 2006.

L.J. Billera and D.C. Heath. Allocation of shared costs: A set of axioms yielding a unique procedure. *Mathematics of Operations Research*, 7(1):32–39, 1982.

R.A. Brealey and S.C. Myers. *Principles of Corporate Finance*. McGraw Hill Series in Finance. McGraw Hill, 7th edition, 2002.

H. Bühlmann. Premium calculation from top down. *ASTIN Bulletin*, 15(2):89–101, 1996.

H. Bühlmann. *Mathematical Methods in Risk Theory*. Springer, Berlin, 1970.

E. Castgnoli, F. Maccheroni, and M. Marinacci. Choquet insurance pricing: A caveat. *Mathematical Finance*, 14(3):481–485, 2004.

M. Denault. Coherent allocation of risk capital. *Journal of Risk*, 4(1):1–34, 2001.

D. Denneberg. *Non-additive Measure and Integral*. Kluwer Academic Publishers, Dordrecht, 1994.

D. Denneberg. Distorted probabilities and insurance premiums. *Methods of Operations Research*, 63:3–5, 1990.

M. Denuit, J. Dhaene, M.J. Goovaerts, and R. Kaas. *Actuarial Theory for Dependent Risk: Measures, Orders and Models*. Wiley, New York, 2005.

M. Denuit, J. Dhaene, M.J. Goovaerts, R. Kaas, and R. Laeven. Risk measurement with equivalent utility principles. *Statistics and Decisions*, 24(1):1–25, 2006.

O. Deprez and H.U. Gerber. On convex principles of premium calculation. *Insurance: Mathematics and Economics*, 4(3):179–189, 1985.

J. Dhaene and M.J. Goovaerts. Dependency of risks and stop-loss order. *ASTIN Bulletin*, 26(2): 201–212, 1996.

J. Dhaene, M. Denuit, M.J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, 31(1):3–33, 2002.

J. Dhaene, A. Tsanakas, E. Valdez, and S. Vanduffel. Optimal capital allocation principles. In *9th International Congress on Insurance: Mathematics and Economics*, Quebec, July 6-8 2005.

P. Embrechts, A. McNeil, and D. Straumann. Correlation and dependence in risk management: Properties and pitfalls. In M.A.H. Dempster, editor, *Risk Management: Value at Risk and Beyond*, pages 176–223, Cambridge, 2002. Cambridge University Press.

European Commission. Considerations on the Design of a Future Prudential Supervisory System. Working paper, MARKT/2535/02EN, Brussels, 2002.

C.J. Exley and A.D. Smith. The cost of capital for financial firms. Technical report, presented to the Institute of Actuaries, 2006.

Financial Services Authority. Integrated Prudential Sourcebook. http://fsahandbook.info/FSA/html/handbook/, 2006.

H. Föllmer and P. Leukert. Quantile hedging. *Finance and Stochastics*, 3(3):251–273, 1999.

H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002a.

H. Föllmer and A. Schied. Robust preferences and convex risk measures. In K. Sandmann and Ph.J. Schönbucher, editors, *Advances in Finance and Stochastics, Essays in Honor of Dieter Sondermann*, pages 39–56, Berlin, 2002b. Springer.

H.U. Gerber. On additive premium calculation principles. *ASTIN Bulletin*, 7(3):215–222, 1974.

H.U. Gerber and M.J. Goovaerts. On the representation of additive principles of premium calculation. *Scandinavian Actuarial Journal*, 4:221–227, 1981.

I. Gilboa and D. Schmeidler. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18:141–153, 1989.

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

M. Goovaerts, R. Kaas, R. Laeven, and Q. Tang. A comonotonic image of independence for additive risk measures. *Insurance: Mathematics and Economics*, 35(3):581–594, 2004.

M.J. Goovaerts, F. De Vylder, and J. Haezendonck. *Insurance Premiums*. Elsevier Science, North-Holland, Amsterdam, 1984.

J. Hancock, P. Huber, and P. Koch. The economics of insurance – How insurers create value for shareholders. Technical report, Swiss Reinsurance Company, Zurich, 2001.

J. Huber. *Robust Statistics*. Wiley, New York, 1981.

S. Kusuoka. On law-invariant coherent risk measures. *Advances in Mathematical Economics*, 3: 83–95, 2001.

R.J.A. Laeven and M.J. Goovaerts. An optimization approach to the dynamic allocation of economic capital. *Insurance: Mathematics and Economics*, 35(2):299–319, 2003.

J. Lemaire. An application of game theory: Cost allocation. *ASTIN Bulletin*, 14(1):61–81, 1984.

H. Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, 1952.

R. Merton and A. Perold. Theory of risk capital in financial firms. *Journal of Applied Corporate Finance*, 6(3):16–32, 1993.

L. Mirman and Y. Tauman. Demand compatible equitable cost sharing prices. *Mathematics of Operations Research*, 7(1):40–56, 2006.

A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks*. Wiley, New York, 2002.

S.C. Myers and J.A. Read. Capital allocation for insurance companies. *Journal of Risk and Insurance*, 68(4):545–580, 2001.

D. Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57 (3):571–587, 2003.

M. Schweizer. Mean-variance hedging for general claims. *Annals of Applied Probability*, 2:171–179, 1992.

D. Tasche. Allocating portfolio economic capital to sub-portfolios. In A. Dev, editor, *Economic Capital: A practitioner's Guide*, pages 275–302, London, 2004. Risk Books.

A. Tsanakas. Dynamic capital allocation with distortion risk measures. *Insurance: Mathematics and Economics*, 35:223–243, 2004.

A. Tsanakas and E. Desli. Risk measures and theories of choice. *British Actuarial Journal*, 9(4): 959–991, 2003.

J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, 1944.

S.S. Wang. Premium calculation by transforming the premium layer density. *ASTIN Bulletin*, 26 (1):71–92, 1996.

S.S. Wang, V.R. Young, and H.H. Panjer. Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics*, 21(2):173–183, 1997.

J.L. Wirch and M.R. Hardy. A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, 25(3):337–347, 1999.

M. Yaari. The dual theory of choice under risk. *Econometrica*, 55(1):95–115, 1987.

V.R. Young and T. Zariphopoulou. Pricing dynamic insurance risks using the principle of equivalent utility. *Scandinavian Actuarial Journal*, 4:246–279, 2002.

CONTRIBUTED TALKS

CONTROL VARIATES FOR CALLABLE LIBOR EXOTICS
A PRELIMINARY STUDY[1]

Jacob Buitelaar[†] **and Roger Lord**[§]

[†]*Equity Derivative Strategies, Goldman Sachs International, River Court, 120 Fleet Street, London EC4A 2QQ, UK*
[§]*Financial Engineering, Rabobank International, P.O. Box 17100, 3500 HG Utrecht, The Netherlands*
Email: `Jacob.Buitelaar@gs.com, roger.lord@rabobank.com`

Abstract

Monte Carlo simulation is currently the method of choice for the pricing of callable derivatives in LIBOR market models. Lately more and more papers are surfacing in which variance reduction methods are applied to the pricing of derivatives with early exercise features. We focus on one of the conceptually easiest variance reduction methods, control variates. The basis of our method is an upper bound of the callable contract in terms of plain vanilla contracts, which is found to be a highly effective control variate. Several examples of callable LIBOR exotics demonstrate the effectiveness and wide applicability of the method.

1. INTRODUCTION

Ever since the seminal papers of Carrière (1996), Tsitsiklis and Roy (2001) and Longstaff and Schwartz (2001) regression methods have become increasingly popular for the valuation and risk management of derivatives with early exercise features. In particular, for high-dimensional non-Markovian models such as the LIBOR market model (LMM), the Longstaff-Schwartz algorithm as it has become known is the method of choice. Though Monte Carlo methods are often criticised for having slow convergence, one distinct advantage over lattice-based methods is that one can appeal to a vast array of probabilistic methods in order to reduce the variance of the estimate of the option price.

Since the introduction of Monte Carlo based methods in mathematical finance, variance reduction techniques, see e.g. Jäckel (2002) and Glasserman (2003) for an extensive overview, have become commonplace when it comes to the valuation of European and path-dependent contracts. It is only recently however that papers have surfaced in which these techniques are applied to

[1]Large parts of this research were carried out while the first author was writing his Master's thesis at the Delft University of Technology and the Modelling and Research department of Rabobank International, and the second author was employed by the latter department and the Tinbergen Institute at the Erasmus University Rotterdam. We thank seminar participants at Rabobank International and the 5th Actuarial and Financial Mathematics Day in Brussels.

the valuation of Bermudan derivatives. Bolia et al. (2004) and Pietersz and Van Regenmortel (2006) consider importance sampling, whereas Piterbarg (2004), Rasmussen (2005), Jensen and Svenstrup (2005), Bolia and Juneja (2005) and Ehrlichman and Henderson (2006) have opted for control variates.

From the numerical examples in these studies it seems that control variates allow for larger orders of magnitude of variance reduction, though specific payoffs such as TARNs are well suited for importance sampling, see Pietersz and Van Regenmortel (2006). Piterbarg (2004) was among the first to suggest using control variates for the valuation of Bermudan derivatives in the LMM. By constructing an analytically tractable Markovian approximation to the LMM it should be possible to value the Bermudan payoff in a lattice and use this as a control variate. This has the advantage that the approach is virtually payoff-independent, the only task left is to construct an approximate model that is highly correlated with the original model.

Other attempts to use control variates have mainly been targeted at using control variates within the original model. Within these papers we can recognise two different types of approaches. On the one hand, Rasmussen (2005) and Jensen and Svenstrup (2005) have opted for using plain vanilla options as control variates, whereas both Bolia and Juneja (2005) and Ehrlichman and Henderson (2006) try to approximate that martingale that renders the additive upper bound of Rogers (2002), Haugh and Kogan (2004) and Andersen and Broadie (2004) equal to the true value of the Bermudan option. Our approach builds on an observation by Jensen and Svenstrup (2005), who noticed that one of the most effective control variates for a Bermudan swaption is simply a cap. First and foremost we aim to provide an intuitive explanation for the effectivity of such simple control variates, and to investigate how they can be improved upon. Second, of the above papers only Jensen and Svenstrup have applied control variates to the pricing of Bermudan interest rate derivatives, and have only considered the easiest example – the Bermudan swaption. The method we consider will be applicable to all kinds of callable LIBOR exotics (CLEs), a term coined by Piterbarg (2004).

This paper is structured as follows. Section 2 introduces some terminology and describes the LMM and CLEs we consider later on. Section 3 reproduces the results that have been reported by Jensen and Svenstrup and analyses why an upper bound on the CLE is a very effective control variate. Section 4 describes how we can easily construct an upper bound for most CLEs and gives two specific examples of this for a callable inverse floater and a snowball. Possible improvements of this idea are also discussed. Finally, section 5 demonstrates the effectiveness of our method and concludes.

2. LIBOR MARKET MODEL AND CALLABLE LIBOR EXOTICS

We start by introducing a tenor structure $T = \{T_i : i = 0, ..., N+1\}$ with daycount fractions over the interval $[T_i, T_{i+1}]$ given by α_i. Next, define the forward LIBOR rate over this time interval as:

$$L_i(t) = \frac{1}{\alpha_i}\left(\frac{P\left(t, T_i\right)}{P\left(t, T_{i+1}\right)} - 1\right)$$

where $P\left(t, T_i\right)$ denotes the time-t price of a zero-coupon bond maturing at T_i. We use the convention that the ith LIBOR sets at T_i and is paid at T_{i+1}. Our numerical results have been generated in the lognormal LIBOR market model in the spot-LIBOR measure. The dynamics of $L_i(t)$ for $0 \leq t \leq T_i$ in this setup are given by:

$$\begin{aligned} dL_i(t) &= \mu_i(t, \boldsymbol{L}(t))L_i(t)dt + \sigma_i(t)L_i(t)dW_i(t) \\ \mu_i\left(t, \boldsymbol{L}(t)\right) &= \sigma_i(t) \sum_{j=m(t)}^{i} \frac{\alpha_j L_j(t)}{1+\alpha_j L_j(t)} \rho_{ij}(t)\sigma_j(t) \\ dW_i(t)dW_j(t) &= \rho_{ij}dt \end{aligned} \tag{1}$$

where σ_i is the volatility of L_i, ρ_{ij} is the instantaneous correlation between L_i and L_j and W_1 through W_N are independent Brownian motions. Finally, $m(t) = \{k \mid T_{k-1} \leq t \leq T_k\}$. In the spot-LIBOR measure the discrete analogue of the money market account is the numeraire asset:

$$B(t) = \frac{P\left(t, T_{m(t)}\right)}{\prod_{j=1}^{m(t)} P\left(T_{j-1}, T_j\right)}.$$

We use the drift approximation of Hunter et al. (2001) to approximate the drift in (1).

Within this model we will be concerned with the valuation of callable or cancellable LIBOR exotics (CLEs). Let us first consider a structured swap, where at a subset[2] T_{pay} of T one exchanges structured coupon payments with floating payments. In particular, for a payer[3] swap the net cashflow payment at T_{i+1} is:

$$\alpha_i \left(F_i - C_i\right). \tag{2}$$

The holder of a receiver swap obviously receives the opposite payments. In our examples the floating rate F_i will be the LIBOR $L_i(T_i)$, though in principle this could be any rate which is known at the payment date T_{i+1}. The most common example for the coupon C_i is $C_i = K$, which amounts to a plain vanilla swap. Other examples we will use are an inverse floater coupon:

$$C_i = \min\left(\max\left(K - L_i(T_i), f\right), c\right), \tag{3}$$

where f indicates a floor and c indicates a cap, and a snowball coupon:

$$C_i = \min\left(\max\left(C_{i-1} + K - L_i(T_i), f\right), c\right). \tag{4}$$

The snowball coupon is similar to the inverse floater, the difference being its path-dependency. In all examples we set $c = \infty$.

A callable version of this swap, or a callable LIBOR exotic, is a Bermudan option to enter into the structured swap at a prespecified set of exercise dates $T_{call} \subseteq T$. Similarly, a cancellable LIBOR exotic gives the right to cancel the swap at a certain set of dates. A parity relationship between callable and cancellable swaps is:

$$\text{Callable payer swap} = \text{Cancellable receiver swap} + \text{payer swap}. \tag{5}$$

When considering options we assume the option holder has the right to call or cancel. If however the option seller has the right to call or cancel, the value to the option holder is exactly the opposite of the value to the option seller. Valuing cancellable options is therefore as easy or difficult as valuing callable options. In the remainder of the paper we use the Longstaff-Schwartz algorithm, see Longstaff and Schwartz (2001), to value the CLEs.

---

[2] In full generality both legs of the swap can have a different frequency and daycount conventions. For ease of exposure we neglect this here.

[3] Analogous to vanilla swaps, a payer swap indicates that we are paying the structured coupon.

3. PRELIMINARY RESULTS

The value of a callable payer swap, where the holder of the swap has the right to call, can be written as an optimal stopping problem:

$$\text{Callable swap}_P(0) = \sup_{\tau \in \tau_{call}} B(0)\mathbb{E}_0 \left[\sum_{i=0}^{\tau} \frac{\alpha_i (F_i - C_i)}{B(T_{i+1})} \right]. \tag{6}$$

Adding a control variate to (6) which is also evaluated at the optimal stopping time t yields:

$$\text{Callable swap}_P(0) = \sup_{\tau \in \tau_{call}} B(0)\mathbb{E}_0 \left[\sum_{i=0}^{\tau} \frac{\alpha_i (F_i - C_i)}{B(T_{i+1})} - \boldsymbol{\beta}^T \left(\boldsymbol{X}(\tau) - \mathbb{E}_0 \left[\boldsymbol{X}(\tau) \right] \right) \right] \tag{7}$$

where $\boldsymbol{X}(t)$ represents a vector of time-t measurable products whose analytical expectation can be evaluated relatively easily. The optional sampling theorem states that the value of (7) is the same as that of (6). In principle we could evaluate the control variate at any other time. Rasmussen (2005) has shown however that it is never optimal to sample the control variate after the optimal exercise date. Though it is not easy to prove that it is never optimal to sample prior to this date, it seems reasonable that the optimal stopping time is also the optimal sampling time of the control variate, as we want the control variates to contain the same information as the payoff. It can be shown that if:

$$\boldsymbol{\beta} = \operatorname{var}\left(\boldsymbol{X}(\tau)\right)^{-1} \operatorname{cov}\left(\sum_{i=0}^{\tau} \frac{\alpha_i (F_i - C)}{B(T_{i+1})}, \boldsymbol{X}(\tau) \right)$$

the variance of (7) is minimised. We will estimate the variance and covariances in this equation by the sample (co)variances in the same simulation. The bias this introduces can be expected to be negligible, see Jäckel (2002).

In this section we will reproduce the results from Jensen and Svenstrup, who found that a cap is a highly effective control variate for a Bermudan swaption. We will try to explain why this is the case. The parameters and settings we use are the ones used by Bender et al. (2005). Tenor dates are chosen as $T_i = 0.5i$, $i = 0, ..., 12\ (= N + 1)$. Daycount fractions are assumed to be constant and equal to 0.5. The volatility functions are time-homogeneous and generated by Rebonato's abcd-formula:

$$\sigma_i(t) = \Phi_i \left(\left[a\,(T_i - t) + d \right] e^{-b(T_i - t)} + c \right). \tag{8}$$

Instantaneous correlations between forward LIBORs are assumed to be:

$$\rho_{ij} = \exp\left(\frac{|j - i|}{N - 2} \ln \rho_\infty \right) \tag{9}$$

for $2 \le i, j \le N$. Values of the parameters in (8) and (9), as well as the initial forward LIBORs are supplied in Table 1. The initial forward term structure is upward sloping, which more often than not seems to be the case in interest rate markets.

i	0	1	2	3	4	5	6	7	8	9	10	11
$L_i(0)$	2.3%	2.5%	2.7%	2.7%	3.1%	3.1%	3.3%	3.4%	3.6%	3.6%	3.6%	3.8%
Φ_i	–	15.3%	14.3%	14.0%	14.0%	13.9%	13.8%	13.7%	13.6%	13.5%	13.4%	13.2%

a	b	c	d	ρ_∞
0.976	2.00	1.500	0.500	0.663

Table 1: Intital forward LIBORs, volatility and correlation parameters

The chosen example in this section is a T_{12} no-call T_1 Bermudan swaption, which allows us to enter into a swap maturing at T_{12} at dates T_1 through T_{11}. In the notation of (2) this implies that the floating rate $F_i = L_i(T_i)$ and $C_i = K$. We will look at an at-the-money (ATM) example, an in-the-money (ITM) and an out-of-the-money (OTM) example. The fixed rates K of these Bermudan swaptions as well as their values are supplied in Table 2. The values in Table 2 have been generated with 100 000 paths. The regressions in the Longstaff-Schwartz algorithm were precomputed in 50 000 independent paths.

	Payer			Receiver		
	K	Value	Se	K	Value	Se
OTM	2.22%	100.6	0.57	4.22%	15.64	0.13
ATM	3.22%	224.2	0.77	3.22%	139.3	0.44
ITM	4.22%	514.1	0.76	2.22%	508.0	0.68

Table 2: Bermudan swaption values and standard errors (in bp)

Let us define the variance reduction ratio by $\kappa = se^2/se_{CV}^2$. Variance reduction is obviously not all we are interested in — if we have found a way to reduce the variance of our price estimate by a factor of 2, but the Monte Carlo simulation including the control variates takes twice as long as the original simulation, the method is not very useful. We take into account the increase in computational time by using the following quantity:

$$\theta = \frac{se^2}{se_{CV}^2} \cdot \frac{\tau}{\tau_{CV}}.$$

The scaling by the ratio of τ to τ_{CV}, the times required for the simulation without and with control variates, accounts for the fact that the computational time required and the inverse of the variance of the price estimate scale roughly linearly with the number of paths. Hence, when e.g. $\theta = 2$ we can obtain the same standard error with control variates in half the time compared to the situation without control variates. In practice we noted that the additional time required to value the control variates within the simulation is minor compared to the total simulation time. The reason for this is that the lion's share of computation time is spent in the construction of the paths. One final note must be made on the computation of θ: here we have neglected the cost of estimating the exercise decision in a separate independent simulation.

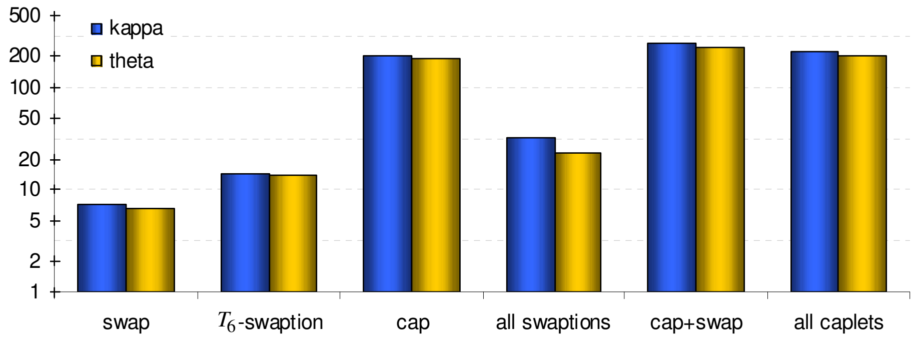

Figure 1: κ and θ for several control variates

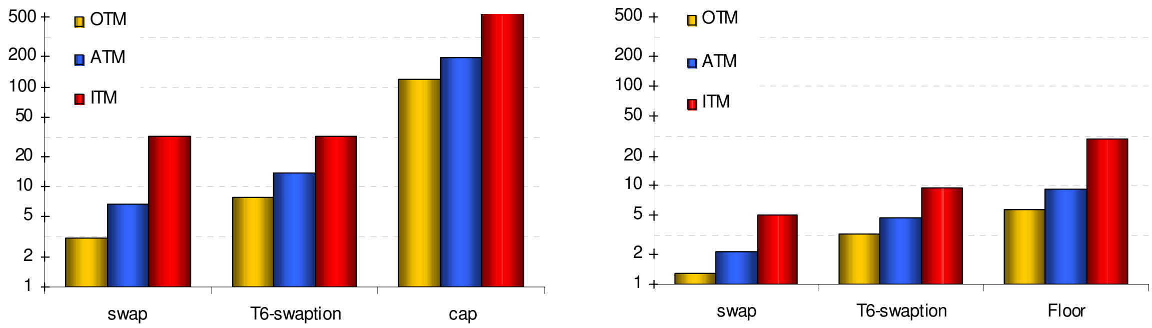

Figure 2: θ for several control variates in OTM/ATM/ITM payer (A) and receiver (B) Bermudan swaption

In Figure 1 we have depicted the variance reduction factor κ and θ, which corrects for the additional computational time that is required, when using several plain vanilla products as control variates for the ATM Bermudan payer swaption from Table 2. The swaptions in Figure 1 are the European swaptions[4] which are embedded in the Bermudan. The reason why only the T_6-swaption is included is that this was found to be the best-performing European swaption.

Even though using a swap or a European swaption shows significant improvements (up to a θ of 15 for the best-performing European swaption), these single control variates are by far outperformed by the cap. The correlation between the cap and the Bermudan is 0.998, i.e. they are almost perfectly correlated. As Jensen and Svenstrup already noted, using a cap alone leads to a higher variance reduction than the best linear combination of all European swaptions which are embedded in the Bermudan swaption. Lifting the restriction on the weights of the caplets (i.e. using a vector of caplets instead of a cap as control variate) does not give large improvements — when investigating the optimal weights of each caplet we noted they were quite close to 1, indicating that the cap is close to optimal. Adding a swap to the cap seems to give the best results, with a 30% improvement compared to the cap alone.

Figure 2 finally shows θ for all options from 2, when a cap — or a floor for the receiver Bermudan — is used as the control variate. As expected, the more ITM the option is, the higher

[4]Theoretically speaking the exact price of a European swaption is not known in closed-form in the LMM. Several highly accurate approximations do exist however, so that the bias induced hereby is negligible.

θ will be as the payoff of the cap or floor will be very highly correlated with the payoff of the underlying swap once we have decided to enter into it. The magnitude of θ is clearly smaller for the receiver Bermudan. What is causing this is actually the same thing that is allowing the cap to work so well: the upward sloping term structure.

The reason why the cap works so well is the following. Once we have decided to enter into the swap, the payments from the swap exactly match the payments of the cap, provided that the floating payment is higher than the fixed. On the optimal exercise moment this is clearly the case. The upward sloping term structure ensures that, on average, the future net exchanges of the payer swap will be positive, so that the value of the cap and the swap will be highly correlated. The initial payments of the cap that occur prior to the exercise date will lower the correlation. However, the contribution of these payments to the cap will most likely be low, as otherwise it is likely that it would have been optimal to exercise into the swap at an earlier date. For the receiver swaption the upward sloping term structure works in exactly the opposite way. After we have decided to exercise, the net value of future payments is expected to be lower than the first payment exchange. As a consequence the probability of negative cash flows increases, leading to a lower correlation of the floor and the structured swap we have exercised into.

Upon inspecting Figure 1 it may seem counterintuitive that the cap works so much better than a single or all embedded European swaptions. A partial explanation for this is that in the extremes (high or low interest rates), the cap perfectly mimics the payoff of the Bermudan swaption. Only the European swaption which has the same start date as the Bermudan shares this property with the cap. An affine combination of control variates will not satisfy this property, which explains what we see in Figure 1.

Even though the swap rate is the main driver behind the exercise decision, a control variate that mimics the payoff of the Bermudan swaption in the extremes will be a highly effective control variate. We further explore this idea for more general CLEs in the next section.

4. EFFECTIVE CONTROL VARIATES FOR CLES

Supported by the results from the previous section, this section will initially focus on generalising the method from the previous section to more general payoffs. In Section 4.2 we try to improve upon this methodology.

4.1. Mimicking the cap

In Section 3 we noted that the cap was a very effective control variate for the Bermudan payer swaption. Observe that the cap is constructed from the underlying swap by taking each forward rate agreement (FRA) and flooring it at zero. This approach is easily extended to other CLEs. The only condition that the payoff has to satisfy is that each floored swaplet can be valued, or at least approximated, in closed-form. We will apply this approach to the two CLEs we presented in Section 2, the inverse floater and the snowball. In the remainder of this text we will refer to a floored swaplet as a “caplet”, and to the sum of all floored swaplets as a “cap”.

Flooring the payoff of the payer inverse floater (see (3) for the coupon) at zero yields:

$$[L_i(T_i) - C_i]^+ = \begin{cases} 2\left[L_i(T_i) - \frac{1}{2}K\right] - [L_i(T_i) - K + f]^+ & f < \frac{1}{2}K \\ -[L_i(T_i) - f]^+ & f \geq \frac{1}{2}K \end{cases}$$

The floored inverse floater coupon is a linear combination of caplets, which can be valued analytically. Hence we can use it as a control variate.

If we consider the payoff of a payer snowball (see (4) for the coupon), it is clear that the underlying swap is the same as that of the inverse floater, with the exception that the strike K has now been replaced by a path-dependent value $C_{i-1} + K$. The easiest way to ensure that the "caplet" can be valued analytically is to replace the path-dependent strike by a constant value, e.g. by assuming that the forward term structure will be realised. Hereafter we can follow the same procedure as for the inverse floater. Alternatively, one could consider a dynamic self-financing strategy where at several points in time during the simulation we update the strikes using all information known at that time. The main drawback hereof is the strong increase in the required computational time. Initial tests of such strategies showed that the increase in computational time required does not outweigh the decrease in variance, so that we omit these results here. Similar findings were reported in Jensen and Svenstrup (2005), where an approximate delta hedge of a Bermudan swaption was considered as a control variate.

4.2. Using multiple control variates

Even though the results in the section 3 were quite impressive, there must be better control variates than the "cap" alone. We can try to improve the variance reduction by altering the strike of the "cap", or even of the individual "caplets". This is indeed confirmed by our investigations, which are not reported here. For the Bermudan swaption, merely changing the strike of the cap/floor increased the factor θ by approximately a factor of 2. In the case of the ATM Bermudan receiver swaption the optimal strike was even as far away as 75 bp from the original strike.

The question is how to determine the strikes that yield the highest correlation with the CLE. In principle one could cache a small number of paths on which we minimise the variance with respect to the free parameters. Subsequently we would generate a larger independent set of paths which we then use to value the Bermudan.

Here we opt for a more pragmatic approach. Focusing on the Bermudan swaption, it is not hard to see that a linear combination of caplets with different strikes can approximate the payoff of a caplet with an unknown strike. In order to enable strikes to vary per maturity, we split the total "cap" into smaller "caps" with different strikes and different caplet maturities. The total amount of control variates must be large enough to have enough flexibility to approximate the unknown optimal control variate, yet small enough to avoid multicollinearity between the control variates.

5. NUMERICAL RESULTS AND CONCLUSIONS

In this section we first investigate whether the general strategy outlined in Section 4.1 yields a satisfactory variance reduction for callable inverse floaters and cancellable snowballs. Second, we investigate whether varying the strike of the suggested control variates is effective, as detailed in Section 4.2.

The exercise or cancellation dates of both products will be the same as that of the Bermudan swaption from Section 3, i.e. we can call or cancel on dates T_1 through T_{11}. Similarly, the cash flow dates are T_2 through T_{12}. For the callable inverse floater we use $K = 6.22\%$ and $f = 2\%$, which ensures that the value of the underlying structured swap is almost zero. For the cancellable snowball we take $C_1 = 1.35\%$ and $K = 3.1\%$. We only consider ATM products in this section, the contract parameters of the inverse floater and the snowball are chosen such that the underlying swap is roughly ATM.

	Bermudan swaption		Callable inverse floater		Cancellable snowball	
	Payer	Receiver	Payer	Receiver	Payer	Receiver
Swap	6.7	2.1	5.9	2.7	n/a	n/a
"Cap"[5]	190.4	8.9	109.8	9.7	7.9	18.8
3 "caps"	370.6	20.3	176.0	23.5	14.3	22.9
3×3 "subcaps"	520.3	62.0	337.8	63.8	26.1	40.5
11×3 "subcaps"	456.0	54.5	n/a	n/a	n/a	n/a

Table 3: θ for various control variates applied to ATM CLEs

Table 3 shows the variance reduction ratio θ we found when using various control variates. The row labelled "swap" reports the results found by using the underlying structured swap as a control variate. The swap underlying the cancellable snowball cannot easily be valued analytically, so that we omit this result from the table. The row labelled "cap" generalises the strategy for the Bermudan swaption to the other CLEs, as outlined in Section 4.1. For the callable inverse floater this clearly leads to a highly effective control variates for the payer case. This confirms our idea that an upper bound of the underlying LIBOR exotic is a highly effective control variate.

For the snowball contract, the magnitude of θ is still quite significant, though not as large as for the other contracts. This is caused by the path-dependency of the snowball coupon. As mentioned earlier, we replaced the previous coupon in (4) by a constant which was calculated by assuming the forward term structure would be realised. Finally, it may seem odd to the reader that the cancellable receiver snowball has a higher θ than the payer contract. This is however caused by the parity relation mentioned in (5).

To investigate whether the suggestions from Section 4.2 are effective, we considered three strategies:

- Use three "caps", where the strikes are equal to K, $K - 1\%$ and $K + 1\%$;

- Split each "cap" into three smaller "subcaps", the first spanning $[T_1, T_3]$, the second spanning $[T_4, T_7]$ and the last spanning $[T_8, T_{11}]$; as control variates we use each of these "subcaps" with strikes equal to K, $K - 1\%$ and $K + 1\%$;

- Split the "cap" into 11 "caplets", and use each of these as a control variates with strikes equal to K, $K - 1\%$ and $K + 1\%$. This leads to a total of 33 control variates.

[5] To be precise, by cap we here mean the payoff which is found by summing all swaplets, floored at zero.

For all CLEs we see that replacing the original cap by a vector of caps gives significantly better results. In this way we manage to get high variance reductions for both the payer and the receiver contracts. Note that the θ found when using a vector of "subcaps" is higher compared to using all "caplets" with three different strikes. The reason for this is that the marginal increase in variance reduction is offset by the increase in computation time. We have therefore only reported these results for the Bermudan swaption, and omitted them for the other CLEs.

Concluding, in this paper we have investigated the result noted by Jensen and Svenstrup that a cap is a highly effective control variate when valuing a Bermudan payer swaption. We have demonstrated how to generalise this idea to more general CLEs by constructing an upper bound for the CLE, and using this as a control variate. This can lead to a large variance reduction, as demonstrated in our numerical results. Benchmarking this method to other control variates, such as the techniques of Bolia and Juneja (2005) and Ehrlichman and Henderson (2006) will be part of further research.

References

L. Andersen and M. Broadie. Primal-dual simulation algorithm for pricing multi-dimensional American options. *Management Science*, 50(9):1222–1234, 2004.

C. Bender, A. Kolodko, and J. Schoenmakers. Iterating snowballs and related path-dependent callables in a multi-factor Libor model. Technical Report 1061, Weierstrass Institute, Berlin, 2005.

N. Bolia and S. Juneja. Function-approximation-based perfect control variates for pricing American options. In M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Joines, editors, *Proceedings of the 2005 Winter Simulation Conference*, pages 1876–1883, Piscataway, New Jersey, 2005. IEEE.

N. Bolia, P. Glasserman, and S. Juneja. Function-approximation-based importance sampling for pricing American options. In R.G. Ingalls, M.D. Rossetti, J.S. Smith, and B.A. Peters, editors, *Proceedings of the 2004 Winter Simulation Conference*, pages 604–611, Piscataway, New Jersey, 2004. IEEE.

J.F. Carrière. Valuation of early-exercise price options using simulations and non-parametric regression. *Insurance: Mathematics and Economics*, 19(1):19–30, 1996.

S.T. Ehrlichman and S.G. Henderson. Adaptive control variates for pricing multi-dimensional American options. Technical report, Cornell University, School of ORIE, 2006.

P. Glasserman. *Monte Carlo methods in financial engineering*. Springer Verlag, New York, 2003.

M. Haugh and L. Kogan. Pricing American options: a duality approach. *Operations Research*, 52 (2):258–270, 2004.

C. Hunter, P. Jäckel, and M. Joshi. Getting the drift. *Risk*, 14:81–84, July 2001.

P Jäckel. *Monte Carlo Methods in Finance*. John Wiley & Sons Ltd., 2002.

M.S. Jensen and M. Svenstrup. Efficient control variates and strategies for Bermudan swaptions in a LIBOR market model. *Journal of Derivatives*, 12(4):20–33, 2005.

F.A. Longstaff and E.S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113–147, 2001.

R. Pietersz and M. Van Regenmortel. Importance sampling for stable Greeks in the LIBOR market model for targeted accrual redemption notes (TARNs). In *Presentation, 3rd Fixed Income Conference*, Amsterdam, 2006.

V. Piterbarg. Pricing and hedging callable LIBOR exotics in forward LIBOR models. *Journal of Computational Finance*, 8(2):65–119, 2004.

N.S. Rasmussen. Control variates for Monte Carlo valuation of American options. *Journal of Computational Finance*, 9(1):2–12, 2005.

L.C.G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12(3):271–286, 2002.

J. Tsitsiklis and B. van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001.

HEDGING UNDER INCOMPLETE INFORMATION
APPLICATIONS TO EMISSIONS MARKETS

Umut Cetin[†] **and Michel Verschuere**[§]

[†]*London School of Economics and Political Science, Department of Statistics, Columbia House, Houghton Street , London WC2A 2AE, United Kingdom*
[§]*Electrabel Trading and Portfolio Management, Regentlaan 8, B-1000 Brussels, Belgium*
Email: u.cetin@lse.ac.uk, michel.verschuere@electrabel.com

Abstract

We study a stochastic model for a market with two tradeable assets where the price of the first asset is implied by the value of the second one and the state of a partially 'hidden' control process. We derive a closed expression for the value of the first asset, as a function of the price for the second and the most recent observation of the control process. We show how the model can be applied to EU markets for carbon emissions.

1. INTRODUCTION TO EU CARBON EMISSIONS MARKETS

The European Commission launched the European Climate Change Programme (ECCP) in June 2000 with the objective to identify, develop and implement the essential elements of an EU strategy to implement the Kyoto Protocol. The Kyoto protocol to the United Nations Framework Convention on Climate Change assigns mandatory targets for the reduction of greenhouse gas emissions to signatory nations. Carbon dioxide, a by-product of the combustion of fossil fuels, is the most widely known type of greenhouse gas. All 25 EU countries simultaneously ratified the Kyoto Protocol on 31 May 2002.

The European Union Emission Trading Scheme (EU ETS) is a significant part of the ECCP and is currently the largest emissions trading scheme in the world. To participate in the EU ETS, members states must first submit a National Allocation Plan (NAP) for approval to the European Commission. Selected carbon intensive installations such as steel manufacturers, power stations of above 20 MW capacity, cement mills, etc. receive free emission credits under the terms of this NAP, enabling them to emit greenhouse gases up to the assigned tonnage.

Installations can bilaterally trade emission certificates under the EU ETS, in order to offset any excess or shortage of carbon emission credits above NAP limits. About 12 000 installations within the Union are covered by the EU ETS in a first phase (2005-2007), representing almost 50% of total carbon emissions. The EU ETS enables selected industries to reduce carbon emissions in a cost effective manner, i.e. installations can opt for either reducing actual carbon emissions or buying additional permits, for instance in case upgrading of the installation would turn out too expensive. The NAP only imposes a cap on the total actual carbon emissions per member state.

Actual trading with EU ETS emission allowances commenced on January 1st, 2005. By the end of the same year, almost 400 million tonnes of carbon equivalent had been traded, representing a turnover in excess of EUR 7 billion. First phase EU ETS carbon credits reached prices of EUR 30 per tonne at the high ending April 2006. Prices for first phase carbon credits then plummeted to below EUR 10 per tonne in a few days beginning May 2006 after EU figures on actual 2005 emission levels suggested emission caps to selected industries had been too generous to have a significant impact. Emission caps for the second phase (2008-2012) are currently under review because of the apparent overdimensioning of NAP levels in the first phase. Further reduction in NAP levels is a likely alternative as well as the inclusion of additional industry sectors such as aviation and transport or still other harmful greenhouse gas emissions like methane.

Selected industries included in the EU ETS can opt to postpone procurement of Phase I emission allowances until Phase II, provided a fine of EUR 40 per tonne is paid at the end of Phase I. This rule actually constitutes a mechanism whereby prices for Phase I and Phase II emission credits become linked in case actual Phase I emissions exceed NAP levels. Prices for Phase I contracts at the end of the first phase will be nonzero only if the EU zone is net short EU ETS carbon credits for this phase. Industries can then opt for borrowing their short ETS position into the next phase at a fixed cost of EUR 40 per tonne.

A net short position in the ETS trading zone thus imposes an identity on prices for Phase I and Phase II emission allowances. In theory, Phase I prices must in that case be equal to Phase II prices plus EUR 40 per tonne. If actual emissions turn out lower than NAP levels, Phase I credits will be worthless at the end of Phase I and the relation with Phase II prices breaks down. Clearly, Phase I emission allowances can be regarded as a derivative with Phase II credits as an underlying, contingent on the net position of the EU ETS zone at the end of 2007. The ‘net position of the zone’ is a variable that cannot be directly observed in the market and hence constitutes a source of non-traded risk that renders the market incomplete.

The present note presents a mathematical model for incomplete markets that can be applied to the context of the EU ETS. We depart from a diffusion model for Phase II prices to which we add an auxiliary process that models the net position of the EU ETS zone on carbon emission allowances. The price of Phase I contracts is conditional on the value of Phase II allowances and the sign of the auxiliary process. Experiences with price behaviour of carbon emission allowances at ending April 2006 has illustrated that the sign, rather than the absolute value of the EU ETS short position can already have a dramatic impact on prices for carbon credits in Phase I. The latter observation suggests a link between Phase I prices and the sign of the short position.

The remainder of this paper is as follows. In Section 2 we briefly summarize the mean-variance approach to the pricing and hedging of contingent claims. We describe a mathematical model in Section 3 for a more general incomplete market and we address the risk neutral pricing of a claim whose value depends on the status of an assumed ‘hidden’ variable. This model is applied to the EU ETS in Section 4 where the price for the Phase I contract price arises, becomes a derivative defined in terms of the auxiliary variable modelling the net position of the EU ETS zone and the Phase II contract price. The price of the Phase I contract takes a canonical form involving the probability of the EU ETS zone being short at the end of Phase I in 2007. Finally, we derive an analytical expression for the Laplace transform of this probability density function, a calculation involving specific properties of an Azema type martingale.

We investigate mathematical properties of this probability density function in Section 4. The results of a number of numerical surveys to the influence of the various parameters are discussed and we present evidence for applicability of the model to actual carbon market data. We finish formulating some conclusions and directives for further research.

2. PRICING AND HEDGING IN INCOMPLETE MARKETS

The issue of pricing and hedging when the market is incomplete is a fairly well studied concept in mathematical finance. As one cannot talk about a single price when the market is no longer complete, there are different methodologies with differing characteristics. One of the most popular methods is to price and hedge using a quadratic risk-minimization criterion. For a short explanation of this approach consider the following example. Let S denote the discounted price process of the risky asset, which is a continuous semimartingale defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ satisfying usual assumptions. Let ξ denote the number of holdings in the risky asset and ν denote the number of holdings in the bond, whose price is normalized to 1. Then the value of the portfolio at time t would be given by $V_t = \xi_t S_t + \nu_t$. Let $C_t := V_t - \int_0^t \xi_s dS_s$. C is the cost process associated with the trading strategy and would be a constant process if the trading strategy were self-financing. Let H be a contingent claim. In a complete market setting we would be able to find a pair (ξ, ν) such that $V_T = H$ and that C were a constant process, i.e. $H = C + \int_0^T \xi_s dS_s$. However, in an incomplete market C would not necessarily be constant since it is typically not possible to find a self-financing replicating strategy. This means one needs to find an optimality criterion to decide on the hedging strategy to be employed. The optimal hedging strategy in this context is defined to be the one that minimizes the remaining quadratic risk at each time t, i.e. $\mathbb{E}[(C_T - C_t)^2] \to \min$. It is shown by several authors (see, e.g., Föllmer and Schweizer (1991), Föllmer and Sondermann (1986), Monat and Stricker (1995) and Schweizer (1991)), under different assumptions that, if one chooses this optimality criterion, the price of the contingent claim is given by

$$X_t = \mathbb{E}^*[H|\mathcal{F}_t],$$

where $\mathbb{E}^*$ corresponds to the expectation operator under the so-called *minimal martingale measure*. The minimal martingale measure, $\mathbb{P}^*$ is an equivalent probability measure under which S is a martingale such that any martingale which was orthogonal to M remains a martingale under $\mathbb{P}^*$ where M is the martingale part in the canonical decomposition of S. Moreover, the optimal hedging strategy $\hat{\xi}$ is given by

$$\xi_t = \frac{d\langle V, S\rangle}{d\langle S\rangle},$$

where $V_t = \mathbb{E}^*[H|\mathcal{F}_t]$. This pricing and hedging methodology is clearly robust under equivalent measure change. We refer the reader to Schweizer (1991), Föllmer and Schweizer (1991), and Monat and Stricker (1995) for further details on the minimal martingale measure and its usage in financial markets.

In the problem that will be analyzed in the next section, the incompleteness arise not only from the fact that there are not enough traded securities to span all the uncertainty in the market but also from the incomplete information. Mathematically speaking, this means the market's information is modelled by the filtration $(\mathcal{G}_t)_{0\le t\le T}$ such that $\mathcal{G}_t \subset \mathcal{F}_t$ for every t where the inclusion is strict. Under this restriction one could define the price of the contingent claim to be

$$P_t = \mathbb{E}^*[H|\mathcal{G}_t].$$

Note that this is the optimal projection of the process X into the filtration $(\mathcal{G}_t)_{0\le t\le T}$ so that the distance between the $\mathcal{F}$-price, X, and $\mathcal{G}$-price, P, is minimized, again, in a quadratic sense. This is the pricing methodology that we will use in the next section.

3. AN INCOMPLETE MARKET MODEL

We consider a market modelled on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P})$ where one of the traded assets $S_t:\ 0 \le t \le T$ follows a dynamics of the Merton type:

$$dS_t = S_t(\mu\, dt + \sigma\, dW_t);\ S_{t=0} = S_0. \tag{1}$$

Denote $\mathcal{F}_t^W:\ 0 \le t \le T$ as the natural filtration for the Wiener process $W_t:\ 0 \le t \le T$, i.e. $\mathcal{F}_t^W = \sigma(W_s:\ 0 \le s \le t)$.

The market is featured of a control process $\theta_t:\ 0 \le t \le T$, described by a Brownian Motion, i.e. we have

$$d\theta_t = d\tilde{W}_t;\ \theta_{t=0} = \theta_0. \tag{2}$$

where $\tilde{W}_t:\ 0 \le t \le T$ is another Wiener process, independent of $W_t:\ 0 \le t \le T$. This SDE has the obvious solution $\theta_t = \theta_0 + \tilde{W}_t:\ 0 \le t \le T$. To ease the exposition of the results we suppose $\theta_0 = 0$, but our result can be easily generalized to the setting where θ_0 is any real number. Define $\mathcal{G}_t:\ 0 \le t \le T$ as the filtration obtained by augmenting the filtration $(\mathcal{F}_t^W)$ by the filtration generated by the signs of θ, i.e $\mathcal{G}_t = \mathcal{F}_t^W \vee \sigma(\mathrm{sign}(\theta_u):\ 0 \le u \le t)$. The function $x \to \mathrm{sign}(x)$ is defined as $\mathrm{sign}(x) = 1$ if $x > 0$ and $\mathrm{sign}(x) = -1$ in case $x \le 0$.

The market trades a contingent claim at price $S_t^0:\ 0 \le t \le T$ entailing the right on the time-T payoff S_T^0 given by

$$S_T^0 = \begin{cases} f(S_T) & \text{if } \theta_T \le 0 \\ 0 & \text{if } \theta_T > 0 \end{cases} \tag{3}$$

where $x \to f(x)$ is bounded real function. The value of the derivative thus depends on the sign of the control process θ_T at expiry. Our goal is to derive the derivative price $S_t^0:\ 0 \le t \le T$, conditioned on the information available from the asset price history $S_t:\ 0 \le t \le T$ and the sign history $\mathrm{sign}(\theta_t):\ 0 \le t \le T$ of the control process θ_t.

To proceed with this calculation, first note that the dynamics (1) for S_t under the minimal martingale measure $\mathbb{P}^*$ becomes a martingale[1] with the dynamics

$$dS_t = S_t\sigma\, dW_t^*;\ S_{t=0} = S_0. \tag{4}$$

[1] $\mathbb{P}^*$ is defined by $\frac{d\mathbb{P}^*}{d\mathbb{P}} = \exp\left(-\frac{\mu}{\sigma}W_T - \frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2 T\right)$.

where $W_t^* : \ 0 \le t \le T$ is a Wiener process under $\mathbb{P}^*$.

The assumption that $\tilde{W}_t$ is independent from W_t implies that S and $\tilde{W}$ are orthogonal martingales with respect to the filtration $\mathcal{G}$ under $\mathbb{P}^*$ by the definition of the minimal martingale measure. Following the lines of argument in the preceding section, the time-t price for the derivative S_T^0 can be computed as

$$S_t^0 = \mathbb{E}^*[S_T^0 \mid \mathcal{G}_t]. \tag{5}$$

We then have

$$\begin{aligned} S_t^0 &= \mathbb{E}^*[S_T^0 \mid \mathcal{G}_t] \\ &= \mathbb{E}^*[f(S_T)I(\theta_T \le 0) \mid \mathcal{G}_t] \\ &= \mathbb{E}^*[f(S_T) \mid \mathcal{G}_t]\mathbb{E}^*[I(\theta_T \le 0) \mid \mathcal{G}_t] \\ &= \mathbb{E}^*[f(S_T) \mid \mathcal{G}_t]\mathbb{P}^*[\theta_T \le 0 \mid \mathcal{G}_t]. \end{aligned}$$

The remaining task is thus to compute the probability $\mathbb{P}^*[\theta_T \le 0 \mid \mathcal{G}_t]$. The filtration $\mathcal{G}_t$ appearing in this expression is of a peculiar type, as it involves both the σ-algebra $\mathcal{F}_t^W$ generated by the price process, S, and the σ-algebra $\sigma(\text{sign}(\theta_u) : 0 \le u \le t)$. Direct computation of this parameter seems a difficult task, and we turn to the Laplace transform of $\theta_T = W_T$ conditioned on $\mathcal{G}_t$ which we write $\mathcal{L}(\theta_T)$. We have

$$\mathcal{L}(\theta_T)(\lambda) = \mathbb{E}^*[\exp(\lambda\theta_T) \mid \mathcal{G}_t] \tag{6}$$

for some $\lambda > 0$. By virtue of the martingale property for exponential Brownian motion, this reduces to

$$\begin{aligned} \mathcal{L}(\theta_T)(\lambda) &= \mathbb{E}^*[\exp(\lambda\theta_T) \mid \mathcal{G}_t] \\ &= \mathbb{E}^*[\exp(\lambda\tilde{W}_T) \mid \mathcal{G}_t] \\ &= \mathbb{E}^*[\exp(\lambda\tilde{W}_T - \frac{\lambda^2 T}{2}) \mid \mathcal{G}_t]\exp(\frac{\lambda^2 T}{2}) \\ &= \mathbb{E}^*[\exp(\lambda\tilde{W}_t - \frac{\lambda^2 t}{2}) \mid \mathcal{G}_t]\exp(\frac{\lambda^2 T}{2}) \\ &= \mathbb{E}^*[\exp(\lambda\tilde{W}_t) \mid \mathcal{G}_t]\exp(\frac{\lambda^2}{2}(T-t)). \end{aligned}$$

We thus need to focus on the computation of $\mathbb{E}^*[\exp(\lambda\tilde{W}_t) \mid \mathcal{G}_t]$ only and this conditioning on $\mathcal{G}_t$ is generally referred to as the Azema martingale. We rewrite

$$\mathbb{E}^*[\exp(\lambda\tilde{W}_t) \mid \mathcal{G}_t] = \mathbb{E}^*[\exp(\lambda|\tilde{W}_t|\text{sign}(\tilde{W}_t)) \mid \mathcal{G}_t].$$

Now define $g_t = \sup\{s < t : \ \tilde{W}_s = 0\}$, i.e. g_t is the last instant preceding time t where the Wiener process $\tilde{W}_t$ passed through the origin. Clearly, g_t is measurable with respect to $\mathcal{G}_t$ and it is known that $M_u := |W_u|/\sqrt{u - g_u}$ obeys the law $\sqrt{2Z}$ for Z exponentially distributed with unit parameter (see Revuz and Yor (1994)). We now use this result to calculate (6). Observe that

$$\mathbb{E}^*[\exp(\lambda\tilde{W}_t) \mid \mathcal{G}_t] = \mathbb{E}^*[\exp(\lambda M_t\,\text{sign}(\tilde{W}_t)\sqrt{t - g_t}) \mid \mathcal{G}_t].$$

This conditional expectation can now be computed, as the law of M_t is known. Let the $\mathcal{G}_t$-measurable random variable A_t be defined as $A_t = \lambda\,\text{sign}(\tilde{W}_t)\sqrt{t - g_t}$, then note that

$$\mathbb{E}^*[\exp(\lambda\tilde{W}_t) \mid \mathcal{G}_t] = \int_0^{+\infty} \exp(A_t\sqrt{2x} - x)\,dx$$

and we make the substitution $u := \sqrt{2x}$ to arrive at

$$\begin{aligned}\mathbb{E}^*[\exp(\lambda\tilde{W}_t) \mid \mathcal{G}_t] &= \int_0^{+\infty} \exp(A_t\sqrt{2x} - x)\,dx \\ &= \int_0^{+\infty} \exp(A_t u - \frac{u^2}{2})\,u\,du \\ &= \int_0^{+\infty} \exp(-\frac{1}{2}(u - A_t)^2 + \frac{A_t^2}{2})\,u\,du\end{aligned}$$

and upon substituting $y := u - A_t$, we finally arrive at

$$\mathbb{E}^*[\exp(\lambda\tilde{W}_t) \mid \mathcal{G}_t] = \int_{-A_t}^{+\infty} \exp(-\frac{1}{2}y^2 + \frac{A_t^2}{2})\,(y + A_t)\,du = 1 + A_t\sqrt{2\pi}\exp(\frac{A_t^2}{2})N(A_t) \quad (7)$$

in which $N(\cdot)$ is the unit normal c.d.f., i.e.

$$N(x) := \int_{-\infty}^{x} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\,dz. \quad (8)$$

This finalizes the computation of the Laplace transform of θ_T, which reads

$$\mathcal{L}(\theta_T)(\lambda) = \exp(\frac{\lambda^2}{2}(T - t))(1 + A_t\sqrt{2\pi}\exp(\frac{A_t^2}{2})N(A_t)) \quad (9)$$

for $A_t = \lambda\,\mathrm{sign}(\tilde{W}_t)\sqrt{t - g_t}$. To summarize, we intend to compute the probability $\mathbb{P}^*[\theta_T \leq 0 \mid \mathcal{G}_t]$ entering (5) via the Laplace transform of θ_T. We only need to invert the Laplace transform in (9) to find the distribution of θ_T given $\mathcal{G}_t$. By integrating this law over the negative half-plane, the desired probability $\mathbb{P}^*[\theta_T \leq 0 \mid \mathcal{G}_t]$ results.

4. APPLICATION TO EU CARBON EMISSIONS MARKETS

The theory out of the previous section is directly applicable to EU carbon emissions markets introduced in Section 1. The traded asset $S_t :\ 0 \leq t \leq T$ in (1) describes the price process for the Phase II emission allowances within the EU ETS in EUR per tonne. The control process $\theta_t :\ 0 \leq t \leq T$ in (2) is interpreted as the net position of the entire EU ETS zone in tonne, relative to NAP levels for all member states added together. Positive values for θ_t indicate that an excess amount of emission certificates is available within the EU ETS perimeter at time t. Alternatively, negative θ_t values are associated to a net short situation in the EU ETS scheme at time t.

The filtration $\mathcal{G}_t$ is the version of the natural filtration $\mathcal{F}_t^W$ for the price process $S_t :\ 0 \leq t \leq T$, augmented with the information contained in the sign of the net EU ETS position $\mathrm{sign}(\theta_t)$ at time t. I.e, $\mathcal{G}_t$ only contains extra information on the net position on EU ETS certificates within the zone, long or short, regardless of the magnitude of the respective excess or shortage. Recent events have indicated that price dynamics within the ETS framework are highly sensitive to market information on just this net position, and the dramatic price-collapse of Phase I emission certificates serves as a direct example. In reality, one will observe the state of the control prices, noteworthy its sign, any time when the EU releases information on the realized carbon emissions or alternatively through estimates provided by selected data providers in the EU carbon market, see PointCarbon (2006) for instance.

The contingent claim $S^0_t : 0 \leq t \leq T$ mimics the price of the Phase I contract and the specific choice $f(x) = x + k$ reflects the regulatory condition that short positions in Phase I allowances can be banked into Phase II at the price of $k = 40$ EUR per tonne as explained in Section 1. The definition for the Phase I payoff in (3) beautifully expresses the true nature of the Phase I contract as a derivative on Phase II prices, conditional on the net position at the end of the first phase, here denoted as $t = T$.

Equation

$$S^0_t = \mathbb{E}^*[f(S_T) \mid \mathcal{G}_t]\mathbb{P}^*[\theta_T \leq 0 \mid \mathcal{G}_t] \tag{10}$$

expresses the value of the Phase I contract in terms of the function f and the probability that the ETS zone will end up short on carbon emission allowances at the end of the first phase $t = T$. After making use of the martingale property for $S_t : 0 \leq t \leq T$ under $\mathbb{P}^*$ and the explicit choice for f, we obtain

$$S^0_t = (S_t + k)\,\mathbb{P}^*[\theta_T \leq 0 \mid \mathcal{G}_t].$$

Note that the probability of the ETS zone ending up short, appearing in the right hand side of (10), can now be expressed as the fraction of Phase I EUA prices divided by the sum of Phase II prices and the penalty of $k = 40$ EUR per tonne, i.e.

$$R_t := \mathbb{P}^*[\theta_T \leq 0 \mid \mathcal{G}_t] = \frac{S^0_t}{(S_t + k)}.$$

The fraction R_t is a known ratio in carbon markets, also referred to as the 'Parson's' ratio PowernextCarbon (2006) and it naturally arises in the present model setting.

We finish by investigating the predictions of the present model for the probability for short position in the ETS zone at the end of Phase I. The Laplace transform (9) can be inverted to yield the conditional probability density for θ_T. Integrating this density over the negative reals results in the desired quantity. Standard inverting of the Laplace transform gives

$$R_t = \mathbb{P}^*[\theta_T \leq 0 \mid \mathcal{G}_t] = \frac{1}{2\pi}\int_{-\infty}^{0} dz \int_{-\infty}^{+\infty} d\lambda \exp(-iz\lambda)\,\mathcal{L}(\theta_T)(i\lambda). \tag{11}$$

This integral can be calculated by numerical methods, resulting in model predictions for the Parson ratio R_t.

We have set $T = 1$ without loss of generality, while t takes values in $[0, T]$ at the same time when g_t satisfies the inequality $0 \leq g_t \leq t$. The next table states selected values for R_t for different values of t and g_t given that the sign of the current position is positive. The t values in the table varies from 0.1 to 0.7, while g_t changes from 0.0 until the t value. Values for R_t were computed using *Mathematica*. The table indicates a 36% probability of the zone ending up short for $t = 0.1$ and $g_t = 0.0$. This probability steadily decreases for larger t values while g_t is kept constant. This is an immediate consequence of the choice we made for the control process: Since a Wiener process is continuous as t approaches T with g_t fixed, $\mathrm{sign}(W_t)$ converges to $\mathrm{sign}(W_T)$. Values for R_t for $t > 0.7$ are more difficult to obtain due to accuracy problems arising in the calculation of (11).

R_t(+)	g_t							
t	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	.34	.50						
0.2	.28	.33	.50					
0.3	.23	.26	.32	.50				
0.4	.18	.21	.25	.31	.50			
0.5	.15	.17	.19	.23	.29	.50		
0.6	.11	.13	.15	.17	.21	.28	.50	
0.7	.08	.09	.10	.12	.15	.18	.25	.50

We have repeated our numerical analysis of the ratio (11) for negative current position and the corresponding table is marked by $(-)$. The conventions are similar as before and results of this effort are summarized in the next table. We note that R_t values now reflect the obvious fact that it is more likely for the zone ending up short if the most recent observation of the control process θ_t pointed in the same direction. Values for the ratio R_t are once more directly implied by the choice for the control as a Wiener process. The table below expresses the fact that the lesser the remaining time in Phase I defined by $T-t$, the less likely the net position of the ETS zone is going to change still. The large values for R_t for $t = 0.7$ are a direct consequence of this.

R_t(-)	g_t							
t	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	.66	.50						
0.2	.72	.67	.50					
0.3	.77	.74	.68	.50				
0.4	.82	.79	.75	.69	.50			
0.5	.85	.83	.81	.77	.71	.50		
0.6	.89	.87	.85	.83	.79	.72	.50	
0.7	.92	.91	.90	.88	.85	.82	.75	.50

The attentive reader will note from the above tables that respective probabilities for $\mathrm{sign}(\theta_t) = \pm 1$ add up to one. This is an immediate consequence of a symmetry present in the Laplace transform of the probability density in (9) as function of the sign of the control process at time t. Characterization of this symmetry is left as an exercise, but eventually boils down to the symmetric nature of marginal distributions for the Wiener process taken as $\theta_t : \ 0 \leq t \leq T$.

5. CONCLUSIONS AND OUTLOOK

We studied a model for an incomplete market with two traded assets. The first asset arises as a derivative on the second and is conditional on the sign of an exogenous control process. The price of the second asset is assumed of the Merton type, while the control process is described by an independent Wiener process. We derive an expression for the price of the first asset under the martingale measure minimizing the quadratic risk induced by the control process. This requires computation of the minimal martingale measure for the prices process of the second asset.

We have shown how our model can be applied to recently established markets for carbon emission allowances and calculated the price for a Phase I contract in terms of the probability of the ETS zone being short at the end of Phase I in December 2007. The probability of for Phase I ETS is referred to as the Parson ratio in most literature and it naturally arises in our model setting.

The price for Phase I emission allowances derived in this paper is strongly dependent on the choice for the underlying control process. Selecting a Wiener process for the control leaves insufficient parameters for calibrating the model to historical data. We currently study a model leaving more degrees of freedom in the control that can be studied in the context of stochastic filtering theory. Results of this survey will be published elsewhere.

Acknowledgement: Michel Verschuere wishes to thank Ingard Moen for useful discussions and proofreading of the manuscript.

References

H. Föllmer and M. Schweizer. Hedging of non-redundant contingent claims under incomplete information. In M.H.A. Davis and R.J. Elliott, editors, *Applied Stochastic Analysis, Stochastics Monographs*, volume 5, pages 389–414. Gordon and Breach, London, 1991.

H. Föllmer and D. Sondermann. Hedging of non-redundant contingent claims. In W. Hildenbrand and A. Mas-Collel, editors, *Contributions to Mathematical Economics in Honor of Gerard Debreu*, pages 147–160. Elsevier Science Ltd, 1986.

P. Monat and C. Stricker. Föllmer-Schweizer decomposition and mean-variance hedging for general claims. *Annals of Probability*, 23(3):605–628, 1995.

PointCarbon, 2006. URL `http://www.pointcarbon.com`.

PowernextCarbon. Tendances carbone, 2006. URL `http://www.caissedesdepots.fr/GB/espace\_presse/publications\_doc/tendances\_carbone\_n5\_eng.pdf`.

D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer Verlag, Berlin, 1994.

M. Schweizer. Option hedging for semimartingales. *Stochastic Processes and Their Application*, 37(2):339–363, 1991.

HEDGING GUARANTEES UNDER INTEREST RATE AND MORTALITY RISK

An Chen[†] and Antje B. Mahayni[§]

[†]*Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands*
[§]*Mercator School of Management, University of Duisburg-Essen, Lotharstr. 65, 47057 Duisburg, Germany*
Email: A.Chen@uva.nl, antje.mahayni@uni-due.de

Abstract

This paper analyzes how model misspecification associated with *both* interest rate *and* mortality risk influences hedging decisions of a life insurer. For this purpose, diverse risk management strategies which are risk-minimizing when model risk is ignored come into consideration. We look at how model risk affects the distribution of the hedging errors associated with these strategies. The analysis is based on endowment assurances which combine an investment element together with a sum assured. Due to periodic premium contributions, i.e., the premium payments stop in the case of an early death, a loan corresponding to the present value of the expected delayed premium payments must be asked for by the insurer in order to implement his hedging decisions. The effect of model risk on this borrowing decision is additionally analyzed.

1. INTRODUCTION

Endowment assurance products are policies which pay out a sum of money on the death of the life assured or at a specified date if the life assured survives the term. This implies that the maturity date and the payoff of the contracts are conditioned on the death time of the life insured.[1] As compensation, the insured provide periodic premiums until the period before the specified date, as long as they are still alive. Obviously, these contracts contain both diversifiable mortality risk and tradeable interest rate risk. Hence, risk management of the issued contracts is based on diversification and hedging, i.e. trading on financial market.

In this paper, we analyze how model risk associated with both the future evolution of the interest rate and the insured's future life expectancy affects the hedging decision of the insurer.

[1]About 75% of the life insurance contracts sold in Germany belong to this category.

I.e. we consider market incompleteness caused by model misspecification associated with both the interest rate and mortality risk. In the analysis of pricing and hedging the risk exposure to the issued contracts, the insurance company makes model assumptions about the term structure of the interest rate and the death distribution. However, the contract fairness and the hedging effectiveness depend on the true interest rate dynamic and the true death distribution. Model misspecification related to the interest rate risk has always been an issue for life insurers because they have difficulties in date inputs to find an appropriate term structure model and reflect the true data generating interest rate process perfectly. Mortality misspecification has attracted more and more attention recently. It can be caused either by a false estimation of the insurer. A medical breakthrough or a catastrophe can lead to an unexpected increase or decrease in life expectancy to a big extent. Moreover, it can result from an intentional abuse of the insurer. For example, an insurer might overestimate the death probability of a potential 70 year's customer deliberately. By doing this, the (assumed) expected period of annuity payment is shortened. Consequently, a higher annuity payment can be offered in order to attract this customer. Due to the fact that it is most of time insurance brokers who close contracts with customers and due to their own interests which might be inconsistent with the insurance company's, this kind of mortality misspecification is not an uncommon phenomenon.

Concerning the literature on model risk, there is an extensive analysis of financial market risk. Without postulating completeness, we refer to the papers of Lyons (1995), Bergman et al. (1996), El Karoui et al. (1998), Hobson (1998), and Mahayni (2003), just to quote a few. Certainly, there are also papers dealing with different scenarios of mortality risk and/or stochastic death distributions, for instance, Milevsky and Promislow (2001), Blake et al. (2006), Ballotta and Haberman (2006), and Gründl et al. (2006). A recent paper of Dahl and Møller (2006) considers the valuation and hedging problems of life insurance contracts when the mortality intensity is affected by some stochastic processes. However, to our knowledge, there are no papers which analyze the distribution of the hedging errors resulting from the combination of both. Therefore, in the present paper, model risk is investigated by studying how it influences the distribution of hedging errors. Speaking of hedging errors, we shall determine the underlying hedging strategies. Neglecting model misspecification, the considered strategies are risk-minimizing. The concept of risk-minimizing is firstly introduced in Föllmer and Sondermann (1986) and applied to the context of insurance contracts in Møller (1998). In the considered contract specification, the most natural hedging instruments are given by the corresponding set of zero coupon bonds. Apparently, a strategy containing the entire term structure is an ideal case. In addition to this ideal case, we also consider a more realistic case where the set of hedging instruments is restricted, i.e. it is only possible to hedge in a subset of bonds.

In order to initialize the above strategies, the insurer needs an amount corresponding to the initial contract value, while he only obtains the first periodic premium at the beginning. Therefore, a credit corresponding to the (assumed) expected discounted value of the delayed periodic premiums should be taken by the insurer, because the initial contract value equals the (assumed) present value of the entire periodic premiums. The insurance company trades with a simple selling strategy to pay back this loan. Apparently, the effectiveness of this strategy in the liability side depends on the model risk too.

It turns out that, independent of the model risk associated with the interest rates, an overestimation of the death probability yields a superhedge in the mean, i.e. the hedger is on the safe side on average. In the case that there is no misspecification with respect to the mortality risk, the model risk concerning the interest rate has no impact on the mean of the hedging error. In contrast, the effect of interest rate misspecification on the variance is crucial, in particular if the set of hedging instruments is restricted. In the case that there is no misspecification with respect to the interest rate dynamic, all strategies considered lead to the same variance level, independent of the mortality. Therefore, the interactivity of both sources of model risk is found to have a pronounced effect on the risk management of the insurer.

The remaining of the paper is organized as follows. Section 2 states the basic features of the insurance contract considered. Section 3 introduces the hedging problem and optimal risk-minimizing hedging strategies. In addition, we study how model misspecification affects the distribution of the hedging errors associated with the relevant strategies. Section 4 illustrates some numerical results for the distributions of the hedging errors under different scenarios of model misspecification. Section 5 concludes the paper.

2. PRODUCT DESCRIPTION

We consider an endowment assurance product with periodic premiums A. In the following, $\underline{\underline{T}} = \{t_0, \ldots, t_{N-1}, t_N\}$ denotes a discrete set of equidistant reference dates where $\Delta t = t_{i+1} - t_i$ gives the distance between two reference dates. The insured pays, as long as he lives, a constant periodic premium A until the last reference date t_{N-1}. In particular, if τ^x denotes the random time of death of a live aged x, then the last premium is due at the random time t_s where $s := \min\{N-1, n^*(\tau^x)\}$ and $n^*(t) := \max\{j \in \mathbb{N}_0 | t_j < t\}$. The insured receives his payoff at the next reference date after his last premium payment, i.e. he receives his payoff at random time $T := \min\left\{t_N, t_{n^*(\tau^x)+1}\right\}$. We denote the endowment part of the contract specification by h and assume that the insured receives at time T the higher amount of h and an insurance account G_T which depends on his paid premiums. Let $\bar{G}_T$ denote the payoff at T, then

$$\bar{G}_T := \max\{h, G_T\}.$$

Notice that the contract specification implies that the benefits and contributions depend on the time of death τ^x. In the case that $G_T = 0$, we have a simple endowment contract which always pays out h amount no matter how the death time of the customer evolves. In particular, the insurance knows exactly its amount of liability but does not know when it is due. In contrast to the simple endowment contract, we consider contracts which also give a nominal capital guarantee, i.e., the insured gets back his paid premiums accrued with an interest rate g ($g \geq 0$), i.e. we use the following convention

$$\tilde{A}_{t_i} := \sum_{j=0}^{i} A e^{g(t_i - t_j)}, \quad G_{t_{i+1}} = \tilde{A}_{t_i} e^{g\Delta t},\ i = 0, 1, \ldots, N-1.$$

To sum up the contract specification, it is convenient to notice that

$$\bar{G}_T = \sum_{i=0}^{N-1} \bar{G}_{t_{i+1}}\ 1_{\{t_i < \tau^x \leq t_{i+1}\}} + \bar{G}_{t_N}\ 1_{\{\tau^x > t_N\}}. \tag{1}$$

3. HEDGING

First, the motivation and derivation of the hedging strategies is based on the value process of the claim to be hedged.

Proposition 3.1 (Value Process) *In our arbitrage-free model setup, the contract value at time $t \in [0, \tau^x]$ is given by*

$$C_t = \sum_{j=n^*(t)+1}^{N-1} \bar{G}_{t_j} \; D(t, t_j) \; {}_{t_{j-1}|t_j}\tilde{q}_{x+t} + \bar{G}_{t_N} \; D(t, t_N) \; {}_{t_{N-1}}\tilde{p}_{x+t},$$

where ${}_{t_{N-1}}\tilde{p}_{x+t}$ denotes the assumed conditional probability that an x-aged life surviving time T given that he has survived t, ${}_{t_{j-1}|t_j}\tilde{q}_{x+t}$ the assumed conditional probability that an x–aged life dies between t_{j-1} and t_j given that he has survived time t, and $D(t, t_j)$ the time t-market price of a zero coupon bond with maturity t_j[2].

Proof. Using Equation (1) and standard theory of pricing by no arbitrage implies that the contract value at t ($0 \leq t < T$) is given by the expected discounted payoff under the martingale measure P^*, i.e.,

$$C_t = E_{P^*}[e^{-\int_t^T r_u du} \bar{G}_T | \mathcal{F}_t].$$

■

The hedging possibility and effectiveness of a claim depend on the set of available hedging instruments. With respect to the insurance contract under consideration, the most natural hedging instruments are given by the set of zero coupon bonds with maturities $t_1, \ldots, t_N$, i.e., by the set $\{D(., t_1), \ldots, D(., t_N)\}$. Thus, we consider the set Φ of hedging strategies which consist of these bonds, i.e.,

$$\Phi = \left\{ \phi = (\phi^{(1)}, \ldots, \phi^{(N)}) \,\middle|\, \phi \text{ is trading strategy with } V(\phi) = \sum_{j=1}^{N} \phi^{(i)} D(., t_i) \right\}.$$

However, due to liquidity constraints in general or transaction costs in particular, it is not possible or convenient to use all bonds for the hedging purpose. This is modelled in the following by restricting the class of strategies Φ. The relevant subset is denoted by $\Psi \subset \Phi$. To simplify the exposition, we propose that the assumed interest rate dynamic is given by a one-factor term structure model and set

$$\Psi = \left\{ \psi \in \Phi \,\middle|\, \psi = (0, \ldots, 0, \psi^{(N-1)}, \psi^{(N)}) \right\}.$$

Two comments are necessary. First, the assumption of a one-factor term structure model implies that two bonds are enough to synthesize any bond with maturity $\{t_1, \ldots, t_N\}$. However, the following discussion can easily be extended to a multi-factor term structure model. Second, as the bonds cease to exist as time goes by, it is simply convenient to use the two bonds with the longest time to maturity.

[2] We put tilde on the death/survival probabilities to denote the assumed ones. Similarly, later we put tilde on the parameters describing term structure of the interest rate to denote the corresponding assumed ones.

As a very conventional hedging criterion used in life insurance contracts, risk-minimizing is applied here as well. Along the lines of Møller (1998), we derive the risk-minimizing hedging strategy for both cases: when the entire term structure or when only the last two zero bonds are used. They are simply denoted by ϕ and ψ respectively. The motivation and derivation of the hedging strategies are based on the value process of the claim to be hedged. Proposition 3.1 immediately motivates a duplication strategy on the set $\{t \leq \tau^x\}$. Prior to the death time τ^x, the contract value (at time t) can be synthesized by a trading strategy which consists of bonds with maturities t_i ($i = n^*(t)+1, \ldots, N$). Assuming that the insurance company will not learn the death of the customer until no further premiums are paid by the insured implies that the strategy proceeds on the set $t \in]\tau^x, T]$ in the same way as on the set $t \in [0, \tau^x]$. Notice that the number of available instruments, i.e. the number of bonds, decreases as time goes by. At time t, only bonds with maturities later than $n^*(t)$ are traded, i.e., the hedger buys $\bar{G}_{t_i} \cdot {}_{t_{i-1}|t_i}\tilde{q}_x$ units of $D(t, t_i)$ and $\bar{G}_{t_N} {}_{t_{N-1}}\tilde{p}_{x+t}$ units of $D(t, t_N)$. The advantage of using this strategy is that the strategy itself is not dependent on the model assumptions of the interest rate.

Proposition 3.2 *Let $\phi \in \Phi$ denote a risk- (variance-)minimizing trading strategy with respect to the set of trading strategies Φ. Assume that the insurance company notices the death of the customer only when no further premium is paid by the insured. If one additionally restricts the set of admissible strategies to the ones which are independent of the term structure, then it holds: ϕ is uniquely determined and for $t \in [0, T]$*

$$\begin{aligned} \phi_t^{(i)} &= 1_{\{t \leq t_i\}} \bar{G}_{t_i} {}_{t_{i-1}|t_i}\tilde{q}_{x+t} \qquad i = 1, \cdots, N-1 \\ \phi_t^{(N)} &= \bar{G}_{t_N} {}_{t_{N-1}}\tilde{p}_{x+t}. \end{aligned}$$

Proof. Without the introduction of model risk it is easily seen that V_{t_0} and the contract value C_{t_0} according to Proposition 3.1 coincide. Thus, with Proposition 3.1 it follows that ϕ is self–financing in the mean. Since the stochastic interest rate risk can be eliminated by trading in all "natural" zero coupon bonds, Møller's (1998) results concerning the independence of mortality and market risk can be adopted here. Since an endowment insurance is a mixture of pure endowment and term insurance, the results immediately follow from Theorem 4.4 and Theorem 4.9 of Møller (1998). ■

A one-factor short rate model is complete in two bonds, i.e. the availability of two bonds with different maturities is enough to synthesize any further bond. Therefore, without postulating the independence from the interest rate model, the variance-minimizing strategy is not defined uniquely.

Proposition 3.3 *Let ψ denote the risk- (variance-)minimizing trading strategy with respect to the set of trading strategies $\Psi \subset \Phi$. Assuming that the insurance company notices the death of the customer only when no further premiums are paid by the insured implies that for $t \in [0, T]$*

$$\begin{aligned} \psi_t^{(N-1)} =& 1_{\{\tau^x \geq t\}} \Bigg(1_{\{t \leq t_{N-2}\}} \sum_{i=n^*(t)+1}^{N-2} \bar{G}_{t_i} {}_{t_{i-1}|t_i}\tilde{q}_{x+t} \frac{D(t, t_i)}{D(t, t_{N-1})} \lambda_1^{(i)}(t) \\ &+ 1_{\{t \leq t_{N-1}\}} \bar{G}_{t_{N-1}} {}_{t_{N-2}|t_{N-1}}\tilde{q}_{x+t} \Bigg) \end{aligned}$$

$$\psi_t^{(N)} = 1_{\{\tau^x \geq t\}} \left(1_{\{t \leq t_{N-2}\}} \sum_{i=n^*(t)+1}^{N-2} \bar{G}_{t_i} \, {}_{t_{i-1}|t_i}\tilde{q}_{x+t} \frac{D(t,t_i)}{D(t,t_N)} \lambda_2^{(i)}(t) \right.$$
$$\left. + \bar{G}_{t_N} \left({}_{t_{N-1}|t_N}\tilde{q}_{x+t} + {}_{t_N}\tilde{p}_{x+t} \right) \right)$$
$$\text{where } \lambda_1^{(i)}(t) := \frac{\tilde{\sigma}_{t_i}(t) - \tilde{\sigma}_{t_N}(t)}{\tilde{\sigma}_{t_{N-1}}(t) - \tilde{\sigma}_{t_N}(t)} \text{ and } \lambda_2^{(i)}(t) := \frac{\tilde{\sigma}_{t_{N-1}}(t) - \tilde{\sigma}_{t_i}(t)}{\tilde{\sigma}_{t_{N-1}}(t) - \tilde{\sigma}_{t_N}(t)}$$

with $\tilde{\sigma}_{\bar{t}}(t)$ denoting the assumed volatility of a zero coupon bond with maturity date $\bar{t}$ at time t.

Proof. Notice that, in the setup of a one-factor short rate model, there is a self-financing strategy $\tilde{\phi}^{(i)} = \left(\alpha^{(i)}, \beta^{(i)}\right)$ with value process $V_t(\tilde{\phi}^{(i)}) = \alpha_t^{(i)} D(t, t_{N-1}) + \beta^{(i)} D(t, t_N) = D(t, t_i)$ for $i = 1, \ldots, N$. One can easily write down the strategy for $D(\cdot, t_i)$, i.e.,

$$\alpha_t^{(i)} = \frac{D(t,t_i)}{D(t,t_{N-1})} \lambda_1^{(i)}(t), \quad \beta_t^{(i)} = \frac{D(t,t_i)}{D(t,t_N)} \lambda_2^{(i)}(t)$$

where $\lambda_1^{(i)}(t)$ and $\lambda_2^{(i)}(t)$ are given as above. Notice that $V_t(\tilde{\phi}^{(i)}) = D(t, t_i)$ P-almost surely implies $\text{Var}[L_T^*(\psi)] = \text{Var}[L_T^*(\phi)]$. This together with $\Psi \subset \Phi$ ends the proof. ■

Obviously, the strategy itself depends on the term structure model of the interest rate. Basically, by using a one-factor interest model, the risk-minimizing strategy for the insurance contract can be implemented in any subset of bonds with at least two elements. A generalization is straightforward if a hedging instrument is added for every dimension of risk factor which is introduced to the short rate model.

It is noticed that the implementation of the above strategies is based on taking a credit at t_0. Since the initial value of the hedging strategies is given by the expected value of the premium inflows, the insurer must in fact borrow the amount $\sum_{i=1}^{N-1} A \, {}_{t_i}\tilde{p}_x \, D(t_0, t_i)$. The underpinning strategy for this is to sell $A \, {}_{t_i}\tilde{p}_x$ bonds with maturity t_i ($i = 1, \ldots, t_{N-1}$). Under mortality risk, it is not necessarily the case that the insurer achieves exactly the number of periodic premiums which are necessary to pay back the credit. These discrepancies lead to extra costs. In particular, these costs can be understood as a sequence of cash flows, i.e., the insurer has to pay back $A \, {}_{t_i}\tilde{p}_x$ at each time t_i ($i = 1, \ldots, t_{N-1}$), i.e. independent of whether the insured survives. Therefore, the additional discounted costs associated with the above borrowing strategy are given by

$$\sum_{i=1}^{N-1} e^{-\int_0^{t_i} r_u \, du} A \left({}_{t_i}\tilde{p}_x - 1_{\{\tau^x > t_i\}} \right). \tag{2}$$

Proposition 3.4 (Expected total discounted hedging costs) *Let L_T^* denote the discounted total costs from both the asset and the liability side. ϕ (ψ) denotes the strategy given in Proposition 3.2 (3.3). Taking account of model risk, $E_{P^*}[L_T^*(\phi)]$ and $E_{P^*}[L_T^*(\psi)]$ own the same value:*

$$D(t_0, t_N) \bar{G}_{t_N} \left({}_{t_N}p_x - {}_{t_N}\tilde{p}_x \right) + \sum_{j=1}^{N-1} \left({}_{t_{j-1}|t_j}q_x - {}_{t_{j-1}|t_j}\tilde{q}_x \right) D(t_0, t_j) \bar{G}_{t_j} + \sum_{i=1}^{N-1} D(t_0, t_i) A \left({}_{t_i}\tilde{p}_x - {}_{t_i}p_x \right).$$

Proof. This proposition is an immediate consequence of Propositions 3.1 ($t = t_0$), in addition to taking the expectation of the addition cost term given in Equation (2). ■

Notice that, independent of the set of bonds, the expected costs are the same. Furthermore, independent of the model risk related to the interest rate, mortality misspecification determines the sign of the expected value, i.e., that decides when a superhedge in the mean can be achieved. When no mortality misspecification is available, the model risk related to the interest rate has no impact on the expected value. When there exists mortality misspecification, the model risk related to the interest rate will influence the size of the expected value. Therefore, the effect of model risk associated with the interest rate depends on the mortality misspecification. However, when it comes to the analysis of the variance, model risk associated with the interest rate has a more pronounced effect than mortality misspecification.

Proposition 3.5 (Additional variance) *It holds*

$$Var_{P^*}[L_T^*(\psi)] \quad = \quad Var_{P^*}[L_T^*(\phi)] + AV_T$$

with $AV_T = 0$ *when there exists no model risk related to the interest rate, otherwise*

$$AV_T = {}_{t_N}p_x \, E_{P^*}\left[(I^*_{t_N}(\psi) - I^*_{t_N}(\phi))^2\right] + \sum_{j=0}^{N-1} {}_{t_j|t_{j+1}}q_x \, E_{P^*}\left[(I^*_{t_{j+1}}(\psi) - I^*_{t_{j+1}}(\phi))^2\right] > 0,$$

where I^* *denotes the discounted gains process, i.e.*

$$I_t^*(\phi) := \sum_{i=1}^{N} \int_0^t \phi_u^{(i)} dD^*(u, t_i).$$

Proof. The proof is given in Chen and Mahayani (2007). ■

It should be emphasized that the effect of mortality misspecification depends on the model risk related to the interest rate. If there exists no interest rate misspecification, mortality misspecification plays no role in the additional variance. However, if there exists model risk related to the interest rate, an additional variance part results always when the restricted subset of zero coupon bonds are used as hedging instruments.

As stated in the introduction, mortality misspecification can be caused by a deliberate use of the insurance company for certain purposes, e.g. safety reasons. I.e., a deviation of the assumed mortality from the true one is generated by a shift in the parameter x which leads to a shift in the life expectancy. For this purpose, we let ${}_tp_{\tilde{x}}$ and ${}_tq_{\tilde{x}}$ denote the assumed probabilities ${}_t\tilde{p}_x$ and ${}_t\tilde{q}_x$.

Proposition 3.6 *For any realistic death/survival probability which satisfies*

$$\frac{\partial \, {}_tp_x}{\partial x} < 0, \text{ and } \frac{\partial_{u|t}q_{x+v}}{\partial x} > 0, \; v \leq u < t,$$

we obtain that

(i) $\frac{\partial E_{P^*}[L_T^*]}{\partial \tilde{x}} < 0$*. Furthermore, an overestimation of the death probability (an underestimation of the survival probability) leads to a superhedge in the mean, i.e.,* $E_{P^*}[L_T^*] \leq 0$*.*

(ii) The additional variance given in Proposition 3.5 is increasing in $\tilde{x}$*.*

Proof. Proof is given in Chen and Mahayani (2007). ■

contract parameter	interest rate parameter (Vasiček model)	mortality parameter (Makeham)
g = 0.05 h = 20 673.6 (G_{t_N} = 35 694.6) t_N = 30 (years) x = 40, A = 500	initial spot rate = 0.05 spot rate volatility = 0.03 speed of mean reversion = 0.18 long run mean = 0.07	 H = 0.0005075787 K = 0.000039342435 c = 1.10291509

Table 1: Basic (assumed) model parameter.

4. ILLUSTRATION OF RESULTS

To illustrate the results of the last sections, we use a one-factor Vasiček-type model framework to describe the financial market risk and a death distribution according to Makeham. The Vasiček-model implies that the volatility $\sigma_{\bar{t}}(t)$ of a zero coupon bond with maturity $\bar{t}$ is $\sigma_{\bar{t}}(t) = \frac{\bar{\sigma}}{\kappa}(1 - \exp\{-\kappa(\bar{t} - t)\})$ where κ and $\bar{\sigma}$ are non-negative parameters. $\bar{\sigma}$ is the volatility of the short rate and κ the speed factor of mean reversion. The death distribution according to Makeham is depicted as follows

$$\begin{aligned} {}_t\tilde{p}_x &= \exp\left\{-\int_0^t \mu_{x+s}\, ds\right\}, \\ \mu_{x+t} &:= H + Kc^{x+t}. \end{aligned} \tag{3}$$

As a benchmark case, we use a parameter constellation along the lines of Delbaen (1990) which is given in Table 1. We introduce the model risk by taking into account that the model parameters which are used to construct the hedging strategies may deviate from the true ones. Concerning the death distribution, we take Makeham hazard rate as an example and concentrate on the mortality misspecifiction caused by the shift in age.

4.1. Expected total costs

Figure 1 demonstrates how the death and survival probability, i.e., ${}_{t_{j-1}|t_j}q_x$ changes with the age x. With the change of x, the death and survival probability demonstrate a parallel shift. If the true age of the customer is 40, then an assumed age of 50 leads to an overestimation of the death probability and an assumed age of 30 results in an underestimation of the death probability. Of course the survival probability and ${}_tp_x$ has exactly a reversed trend.

How the expected discounted total costs from both asset and liability side change with the assumed age $\tilde{x}$ is depicted in Figure 2. It is noticed that, for the given parameters, the expected discounted total cost exhibits a negative relation in $\tilde{x}$. The higher $\tilde{x}$, the lower the expected total costs. It is observed that, independent of the set of hedging instruments (bonds), the hedger achieves profits in mean (negative expected discounted cost) if he overestimates the death probabilities.[3] Hence, negative expected discounted costs result when true x is smaller than the assumed

[3] This result is opposite to the result in pure endowment insurance contracts, where a negative expected discounted cost is achieved when an overestimation of the survival probability exists.

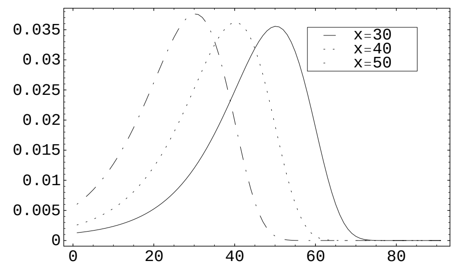

Figure 1: ${}_{t_{j-1}|t_j}q_x$ for $x = 30, 40, 50$. The other parameters are given in Table 1.

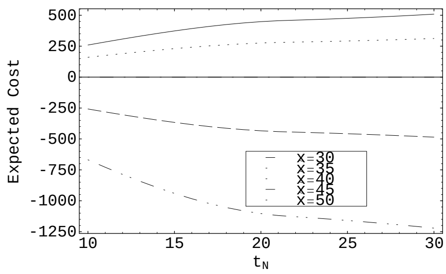

Figure 2: Expected cost for $\tilde{x} = 30, 35, 40, 45, 50$ with the real $x = 40$. The other parameters are given in Table 1.

one. Converse effects are observed when the insurer underestimates the death probability. Here, a real age of 40 is taken and it is observed that for $\tilde{x} = 45, 50$, the expected costs have negative values (lower two curves), and for $\tilde{x} = 30, 35$, the expected costs exhibit positive values. When the true age coincides with the assumed one, the considered strategy is mean-self-financing because the expected discounted cost equals zero.

4.2. Variance Difference

In contrast to the expected total costs, the distribution of the costs depends on the set of hedging instruments. This subsection attempts to illustrate how the variance difference depends on the model risk. Assuming that the short rate is driven by a one-factor Vasiček model, model risk associated with the interest rate can be characterized either by the mismatch of the volatility ($\bar{\sigma}$) or the speed factor (κ), which are determining factors in the volatility function of the zero coupon bonds. Due to the Vasiček modelling, the misspecification of $\bar{\sigma}$ has no impact on the variance difference. Therefore, in the following, we concentrate on the interest rate misspecification characterized by the deviation of the assumed $\tilde{\kappa}$ from the true κ.

We obtain some values for the variance difference as exhibited in Table 2. Firstly, there exists a deviation of $\tilde{\kappa}$ from κ, the variances of these two strategies differ, even when there is no mortality misspecification. Secondly, mortality misspecification does not have impact on the variance difference, if there are no interest rate misspecification available. I.e., these two strategies make no difference to the variance of the total cost if no model risk associated with the interest rate appears. Therefore, for $\tilde{\kappa} = \kappa = 0.18$, overall the variance difference exhibits a value of 0. These two observations validate the argument that the model misspecification resulting from the term structure of the interest rate has a substantial effect when the variance is taken into account. The effect of mortality risk is partly contingent on the model risk associated with the interest rate. Thirdly, only the absolute distance of $\tilde{\kappa}$ from κ counts. The bigger this absolute distance is, the higher variance differences these two strategies result in. Therefore, overall you observe parabolic curves for the variance difference. In addition, the variance difference increases in $\tilde{x}$. This positive effect can be observed in Figures 3 and 4.

To sum up, if the hedger substantially overestimates ($\tilde{\kappa} << \kappa$) or underestimates ($\tilde{\kappa} >> \kappa$) the

	Expected total cost			Variance Difference			The Ratio		
$\tilde{\kappa}$	$\tilde{x}=35$	$\tilde{x}=40$	$\tilde{x}=45$	$\tilde{x}=35$	$\tilde{x}=40$	$\tilde{x}=45$	$\tilde{x}=35$	$\tilde{x}=40$	$\tilde{x}=45$
0.150	287.795	0	−449.842	774.983	1876.71	4660.65	0.0967	–	−0.1518
0.155	293.229	0	−457.831	568.654	1375.73	3414.78	0.0813	–	−0.1276
0.160	297.998	0	−464.843	385.079	930.677	2308.87	0.0659	–	−0.1034
0.165	302.192	0	−471.014	229.521	554.135	1373.97	0.0501	–	−0.0787
0.170	305.891	0	−476.458	108.254	261.073	646.952	0.0340	–	−0.0534
0.175	309.159	0	−481.270	28.7651	69.2933	171.608	0.0173	–	−0.0272
0.180	312.054	0	−485.532	0	0	0	0	–	0
0.185	314.621	0	−489.315	32.6569	78.4791	194.106	0.0182	–	−0.0285
0.190	316.903	0	−492.677	139.546	334.922	827.806	0.0373	–	−0.0584
0.195	318.934	0	−495.670	336.029	805.425	1989.29	0.0575	–	−0.0890
0.200	320.745	0	−498.338	640.547	1533.21	3783.96	0.0789	–	−0.1234
0.205	322.361	0	−500.719	1075.28	2570.10	6338.04	0.1017	–	−0.1590
0.210	323.805	0	−502.847	1666.92	3978.35	9802.85	0.1261	–	−0.1969

Table 2: Expected total cost, variance differences and the ratio of the standard deviation of the variance difference and the expected total cost for varying $\tilde{\kappa}$ with $x = 40$ and the other parameters are given in Table 1.

bond volatilities, and if at the same time he highly overestimates the death probability ($\tilde{x} >> x$), the diverse choice of the hedging instruments leads to a huge difference in the variance. On the contrary, a $\tilde{\kappa}$ value close to κ combined with a big overestimation of the survival probability ($\tilde{x} << x$) almost leads to very small variance difference. I.e., very close variances result. The choice of the hedging instrument does not have a significant effect under this circumstance. These result leads to a very interesting phenomenon, with an overestimation of the death probability ($\tilde{x} > x$), the insurance company is always on the safe side in mean, i.e., it achieves a superhedge in the mean. However, if the set of hedging instruments is restricted, an overestimation of the death probability does not necessarily decrease the shortfall probability under a huge misspecification associated with the interest rate (characterized by a big deviation of $\tilde{\kappa}$ from κ). This is due to the

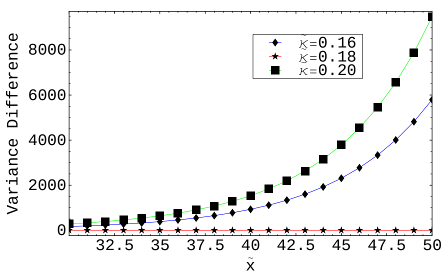

Figure 3: Variance difference as function of $\tilde{x}$ with the real $x = 40$ for $\tilde{\kappa} = 0.16$, $\tilde{\kappa} = 0.18$ and $\tilde{\kappa} = 0.20$. The other parameters are given in Table 1.

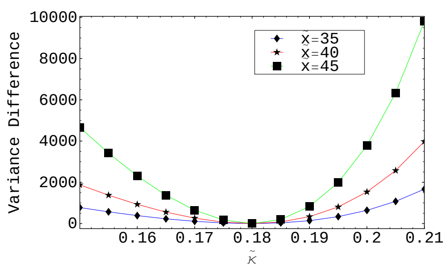

Figure 4: Variance difference as function of $\tilde{\kappa}$ with the real $\kappa = 0.18$ for $\tilde{x} = 35$, $\tilde{x} = 40$ and $\tilde{x} = 45$. The other parameters are given in Table 1.

observation that a quite high variance difference is reached under this parameter constellation.

In addition, due to the tradeoff between the expected value and the variance difference[4], it is interesting to have a look at the relative size, like the ratio of the standard deviation of the variance difference and the expected value of the total cost from both asset and liability side. First of all, this ratio is not defined when the assumed and real age coincide. Second of all, here for the given parameters, an overestimation of the death probability ($\tilde{x} = 45$) has a higher effect than an underestimation ($\tilde{x} = 35$), i.e. the absolute value of this ratio is larger for the case of $\tilde{x} = 45$. Finally, this ratio can give a hint to the safety loading factor. Assume, the insurer uses standard–deviation premium principle. The ratio given in Table 2 suggests him how much safety loading to take when he uses the last two bonds instead of the entire term structure.

5. CONCLUSION

The risk management of an insurance company must take into account model risk, i.e. the uncertainty about the interest rate and the life expectancy. We show that even a small difference between assumed and realized death scenarios may have a great impact on the hedging performance because of the existence of interest rate risk. In practice, this is particularly important because a deviation of true and assumed mortality/survival probabilities is unavoidable and sometimes even caused intentionally by the insurance company itself. The problem which is associated with the interdependence of model risk concerning the interest rate dynamic and the mortality distribution is even more severe if there is a restriction on the set of hedging instruments. We measure the risk implied by the restriction of hedging instruments by calculating the additional variance of the hedging costs, i.e. the variance which is to be added to the variance term without the restriction. Further, we stress an important problem which arises if, as it is normally done, the contributions of the insured are given in terms of periodic premiums instead of an up-front premium. If the contributions of the insured are delayed to a future, uncertain time, model risk influences the liability side in addition to the asset side. Theoretically, a credit must be taken by the insurer in order to implement the considered hedging strategies in the asset side. The insurer achieves not necessarily the number of periodic premiums which is needed to pay back his credit, which leads to an extra cost to the insurer. To sum up, neither the model risk which is related to the death distribution nor the one associated with the financial market model is negligible for a meaningful risk management.

References

L. Ballotta and S. Haberman. The fair valuation problem of guaranteed annuity options: The stochastic mortality environment case. *Insurance: Mathematics and Economics*, 38(1):195–214, 2006.

[4]An overestimation of the death probability leads to a superhedge in the mean but at the same time a higher variance difference.

Y.Z. Bergman, B.D. Grundy, and Z. Wiener. General properties of option prices. *Journal of Finance*, 51(5):1573–1610, 1996.

D. Blake, A.J.G. Cairns, and K. Dowd. Pricing death: Frameworks for the valuation and securization of mortality risk. *ASTIN Bulletin*, 36:79–120, 2006.

A. Chen and A.B. Mahayani. Hedging endowment assurance products under interest rate and mortality risk. Technical report, Bonn Graduate School of Economics, University of Bonn, 2007.

M. Dahl and T. Møller. Valuation and hedging of life insurance liabilities with systematic moratility risk. *Insurance: Mathematics and Economics*, 39:193–217, 2006.

N. El Karoui, M. Jeanblanc-Picqué, and S.E. Shreve. Robustness of the Black and Scholes formula. *Mathematical Finance*, 8(2):93–126, April 1998.

H. Föllmer and D. Sondermann. Hedging of non-redundant contingent claims. In W. Hildenbrand and A. Mas-Collel, editors, *Contributions to Mathematical Economics in Honor of Gerard Debreu*, pages 205–233, Amsterdam, 1986. North Holland.

H. Gründl, T. Post, and R.N. Schulze. To hedge or not to hedge: Managing demographic risk in life insurance companies. *Journal of Financial and Quantitative Analysis*, 73(1):19–41, 2006.

D. Hobson. Volatility misspecification, option pricing and superreplication via coupling. *Annals of Applied Probability*, 8(1):193–205, 1998.

T.J. Lyons. Uncertain volatility and the risk-free synthesis of derivatives. *Applied Mathematical Finance*, 2:117–133, 1995.

A. Mahayni. Effectiveness of hedging strategies under model misspecification and trading restrictions. *International Journal of Theoretical and Applied Finance*, 6:521–552, 2003.

M.A. Milevsky and S.D. Promislow. Mortality derivatives and the option to annuitise. *Insurance: Mathematics and Economics*, 29:299–318, 2001.

T. Møller. Risk-minimizing hedging strategies for unit-linked life insurance contracts. *Astin Bulletin*, 28:17–47, 1998.

DEALING WITH THE VOLATILITY SMILE OF HIMALAYAN OPTIONS

Joseph Meaney

School of Mathematical Sciences, University College Cork, Ireland
Email: joseph.meaney@gmail.com

Abstract

The risks involved in the Himalaya options market have never been more evident than in recent years as these options have been responsible for heavy losses over that period. These losses have occurred not only because of adverse market conditions but also because of model misspecifications. Principal among the dangers of pricing is the inability to find a volatility model which is consistent with the observed volatility surfaces of individual stocks involved in the Himalayan. In this paper we present a solution to this problem by using a multivariate mixture of densities technique. We outline the practical implications of extending this technique to the Himalayan setting and present our numerical results in detail.

1. INTRODUCTION

Himalayan options first appeared in 1998 when they were introduced by Société Générale as part of their mountain range series of options. At the time the Himalayan was an entirely new type of derivative product that merged the path dependency of barrier options with the multi dimensionality of basket options. Since then several derivatives have emulated the structure of the Himalayan most notably the Emerald option which has been introduced on Nordic markets.

Although the Himalayan contains some very exotic traits they can be classified as European in nature and as such their pricing is heavily dependent on the usual sensitivities, namely correlation estimation and the volatility smile. In any basket type option correlation plays a major role in pricing, thus significant care is required when estimating this parameter. Including historical correlation directly in pricing is a risk since underlying correlations can change significantly during times of market stress. The issue of modeling the volatility smile is of great practical importance for a Himalayan because when stocks are removed from the basket the smile can become distorted.

The primary concern of this paper is to apply a mixture of densities technique to the pricing of Himalayan options in order to deal with the volatility smile. This is done by applying a method developed by Brigo et al. (2004) which constructs an entirely new manifold of local volatility structures and removes the need for costly time discretization which is common for the conventional mixture of densities model.

This paper is split into four sections. The first is a brief review of the Himalayan option including a description of some of the different variants available. The second covers the conventional approach to the multidimensional mixture of densities and then explains the multivariate mixture of densities (MVMD) developed by Brigo et al. (2004). We then apply their technique to the Himalayan option and finally we end with conclusions and thoughts on further work.

2. HIMALAYAN OPTIONS

A Himalayan may most simply be described as a call on the sum of the best performers on a basket of stocks over a certain time interval. At the beginning of the contract a collection of exercise dates is decided upon. The number of dates may be equal to the number of stocks in the basket or depending on the variant of the Himalayan it could be less. At each of these dates the best performing stock (according to a specified performance measure) is permanently removed from the basket and its return becomes a coupon. This process is then repeated at each exercise date until there is only one stock left.

Suppose we have a Himalayan written on the underlying stocks $S_1, S_2, \ldots, S_n$ and that the contract is active over the time interval $[0, T]$. This interval is then split into n subintervals written as:

$$[t_0, t_1], [t_1, t_2], \ldots, [t_{i-1}, t_i], \ldots, [t_{n-1}, t_n]$$

where $t_0 = 0$ and $t_n = T$. The end point of each of these intervals represents the n exercise dates where the best performing stock will be removed from the basket. Ignoring discounting, the payoff of a Himalayan will then take one of the following forms:

$$\max\left\{A\sum_{i=1}^{n}\left(\frac{S_{m(i)}(t_i)}{S_{m(i)}(0)} - 1\right), 0\right\} \tag{1}$$

or

$$A\sum_{i=1}^{n}\max\left\{\left(\frac{S_{m(i)}(t_i)}{S_{m(i)}(0)} - 1\right), 0\right\} \tag{2}$$

where A is the nominal amount associated with the contract and the index $m(i)$ records the best performer of the stocks which remain on the interval $[t_{i-1}, t_i]$. We will illustrate the difference between both variants by way of an example.

	Months						
	0	1	2	3	4	5	6
A	105.18	100.00	96.54	87.45	91.40	92.89	89.06
B	9.12	10.63	11.14	15.67	14.12	13.56	16.76
C	48.97	50.90	58.97	55.40	60.20	61.23	65.45
D	15.62	17.53	25.73	30.91	32.00	34.71	38.49
E	90.14	100.05	105.36	107.86	111.65	116.23	121.49
F	90.36	84.99	89.11	93.06	96.96	99.01	94.62

Table 1: Stock Price Table

	Monthly Returns						
	1	2	3	4	5	6	Coupon
A	−0.05	−0.08	−0.17	−0.13	−0.12	<u>−0.15</u>	−0.15
B	<u>0.16</u>						0.16
C	0.04	0.20	0.13	<u>0.23</u>			0.23
D	0.12	<u>0.64</u>					0.64
E	0.11	0.17	<u>0.20</u>				0.20
F	−0.06	−0.01	0.03	0.07	<u>0.10</u>		0.10

Table 2: Stock Price Returns

Suppose we have a Himalayan of the type (1) written on six underlying stocks A, B, C, D, E and F with table 1 charting the closing price of these stocks over six time periods. Table 2 calculates the terminal payoff of the Himalayan in this situation. The first column of table 2 shows the return of each stock over period one. As is shown B has the highest return thus this return is locked in as the first coupon and B is removed from the basket. This process is then repeated in the next 5 columns with the final column showing the contribution of each stock to payoff of the Himalayan. The payoff is then the sum of the coupons. In this case the payoff was 1.18. The Himalayan described in (2) could have improved on this payoff since the contribution of A would have been floored at zero giving a payoff of 1.33.

Another variation to the Himalayan is to payout on fewer periods:

$$\max\left\{A\sum_{i=1}^{L}\left(\frac{S_{m(i)}(t_i)}{S_{m(i)}(0)}-1\right),0\right\} \tag{3}$$

or

$$A\sum_{i=1}^{L}\max\left\{\left(\frac{S_{m(i)}(t_i)}{S_{m(i)}(0)}-1\right),0\right\} \tag{4}$$

where $L < n$. This can have the effect of increasing the payout since stocks that have performed badly over the entire contract can be excluded from the payout. In our example if we had let $L = 5$ then the payout would have excluded A which performed poorly.

The structure of the Himalayan can be changed drastically by calculating the returns in a different manner. Instead of returns being relative to the initial prices at the start of the contract we can calculate returns over a period relative to the opening price of stocks at the beginning of that period. This leads to the following pricing functions:

$$\max\left\{A\sum_{i=1}^{n}\left(\frac{S_{m(i)}(t_i)}{S_{m(i)}(t_{i-1})}-1\right),0\right\} \tag{5}$$

$$A\sum_{i=1}^{n}\max\left\{\left(\frac{S_{m(i)}(t_i)}{S_{m(i)}(t_{i-1})}-1\right),0\right\} \tag{6}$$

where again $m(i)$ indexes the best performing stock from those that remain over the period $[t_{i-1}, t_i]$. This variation tends to wipe out the bad past performances of stocks and although returns will decrease as more stocks are removed from the basket the final payout will in general be higher than the structure (1) and (2). For a more detailed description of the Himalayan option see Overhaus (2002).

3. MIXTURE OF DENSITIES (MD)

The issue of volatility modeling is of great importance for Himalayan options since each of the stocks involved with the option and the basket itself are likely to show a volatility smile. In order to deal with this we propose to use a multivariate mixture of densities (MVMD) technique developed by Brigo et al. (2004) in the context of Himalayas.

Suppose we have a stock S which exhibits a non constant volatility and obeys dynamics given by the equation

$$dS_t = \mu S_t dt + \sigma(S_t, t)S_t dW_t$$

under the T forward measure Q^T where μ is constant, W is a standard Brownian motion and σ is well behaved function which obeys linear growth conditions. Under these circumstances a unique solution to the above SDE exists. The aim of the MD technique is to find the local volatility of S_t in terms of the marginal densities of N auxiliary processes z_t^k. These are governed by the dynamics

$$dz_t^k = \mu z_t^k dt + v_k(z_t^k, t)dW_t$$

where each z_t^k has marginal density p_t^k and $v_k(z_t^k, t)$ satisfies linear growth conditions. The marginal density of S_t is then assumed to be representable as a convex combination of the marginal densities of z_t^k,

$$p_t(S) = \sum_k \lambda_k p_t^k$$

with $\lambda_k \geq 0$ for all k and $\sum_k \lambda_k = 1$. Through manipulations of the related forward Kolmogorov equations it can be shown that a candidate for the local volatility of S_t can be written as

$$\sigma(y,t) = \sqrt{\frac{\sum_{k=1}^{N} \lambda_k v_k^2(y,t) p_t^k}{\sum_{k=1}^{N} \lambda_k y^2 p_t^k}}. \tag{7}$$

If the triplet (z_t^k, v_k, p_t^k) is defined as

$$z_0^k = S_0 \quad v_k(y,t) = y\sigma_k(t) \quad V_k(t) = \sqrt{\int_0^t \sigma_k^2(s)ds}$$

and

$$p_t^k(y) = \frac{1}{\sqrt{2\pi} y V_k(t)} \exp\left[-\frac{1}{2V_k^2(t)} \left(\log\left(\frac{y}{S_0}\right) - \mu t + \frac{1}{2} V_k^2(t) \right)^2 \right]$$

the corresponding dynamics for $S(t)$ admits a unique strong solution with local volatility as shown previously. As a consequence of (7) pricing a European option becomes immediate since mixing basic lognormal densities leads to analytically tractable option prices under the model process for the underlying asset. In the case of a European call on the stock S with maturity T and strike K the initial price is

$$D_{0T} E[S_T - K]^+ = \sum_{i=1}^{N} \lambda_i D_{0T} \int_0^{+\infty} [x-K]^+ p_T^i(x) dx$$

where $p_T^i(x)$ are the marginal densities of the auxiliary equations and D_{0T} is the discounting factor over $[0,T]$. This MD model allows accurate calibration to any smile shaped volatility curve or surface. Freedom when choosing the number of auxiliary equations offers great flexibility and as a consequence of the pricing formula for European calls calibration to entire implied volatility structures over several strike prices and maturities are possible. We now leave the single stock MD and look at the multivariate case.

Suppose we wish to evaluate the price of an option which depends on several underlying assets that each show a volatility smile. In order to price an option such as this, using the conventional MD method, we would begin by calibrating the volatility surface of each of the stocks using the single asset MD shown in the previous section. Once the dynamics of each of the stocks is found a Monte Carlo simulation would be used to price the option by using paths suitably discretized according to the drift rate of each of the stocks and a diffusion matrix given by the local volatility. Using an instantaneous correlation structure ρ_{ij} calculated through a technique such as historical analysis and supposed constant over time, the Monte Carlo method would proceed by simulating the joint evolution of the stocks over the time grid $[\tau_0, \tau_1], \ldots, [\tau_{n-1}, \tau_n]$ with a covariance matrix component (i,j) over the time interval $[\tau_m, \tau_{m+1}]$ given by:

$$C(S_i, S_j, t) = \sqrt{\frac{\sum_{k=1}^{N} \lambda_{ik} \sigma_{ik}^2(t) p_{\tau_m}^{ik}(S_i)}{\sum_{k=1}^{N} \lambda_{ik} p_{\tau_m}^{ik}(S_i)}} \sqrt{\frac{\sum_{k=1}^{N} \lambda_{jk} \sigma_{jk}^2(t) p_{\tau_m}^{jk}(S_j)}{\sum_{k=1}^{N} \lambda_{jk} p_{\tau_m}^{jk}(S_j)}} \rho_{ij}. \tag{8}$$

Thus the stocks in our basket will evolve according to

$$S_i(\tau_{m+1}) = S_i(\tau_m) + \mu S_i(\tau_m)(\tau_{m+1} - \tau_m) + \sqrt{\frac{\sum_{k=1}^{N} \lambda_{ik} \sigma_{ik}^2(\tau_m) p_{\tau_m}^{ik}(S_i)}{\sum_{k=1}^{N} \lambda_{ik} p_{\tau_m}^{ik}(S_i)}} S_i(\tau_m)(W_{\tau_{m+1}} - W_{\tau_m}).$$

This scheme will certainly price an option on a basket of stocks. However, by imposing that the covariance of the multidimensional process to be of the form of (8) we will be moving within a given manifold of the possible local volatility structures. Discretization is also a costly procedure for even the most simple of derivatives.

The multivariate mixture of densities (MVMD) technique of Brigo et al. (2004) deals with these problems by forming the marginal density of the multivariate process in a completely new fashion. Again suppose we wish to price an option written on the underlying assets $S_1, \ldots, S_n$ with marginal densities p_t^i and that we have calibrated a MD to each of their volatility surfaces according to:

$$p_t^i(y) = \sum_{k=1}^{N_i} \lambda_{ik} p_t^{ik}(y), \text{ with } \lambda_{ik} \geq 0, \forall k \text{ and } \sum_{k=1}^{N_i} \lambda_{ik} = 1$$

where each S_i has N_i auxiliary equations given by

$$dz_t^{ik} = \mu_{ik} z_t^{ik} dt + \sigma_{ik}(t) z_t^{ik} dW_t.$$

For simplicity we'll assume that each stock has the same number of auxiliary equations N. The MVMD then proceeds as follows. The joint multivariate density is defined as

$$p_t(\mathbf{y}) = \sum_{i_1, i_2, \ldots, i_n = 1}^{N} \lambda_{1i_1} \lambda_{2i_2} \ldots \lambda_{ni_n} p_t^{(i_1 \ldots i_n)}(\mathbf{y})$$

where the component densities are

$$p_t^{(i_1 \ldots i_n)}(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Xi^{(i_1 \ldots i_n)}(t)} \prod_{i=1}^n y_i} \exp\left[-\frac{\tilde{\mathbf{y}}^{(i_1 \ldots i_n)} \Xi^{(i_1 \ldots i_n)}(t)^{-1} \tilde{\mathbf{y}}^{(i_1 \ldots i_n)}}{2} \right]$$

with

$$\tilde{y_l}^{(i_1 \ldots i_n)} = \ln y_l - \ln y_l(0) - \int_0^t \left(\mu_s^{(li_l)} - \frac{\sigma_s^{(li_l)^2}}{2} \right) ds$$

and $\Xi^{(i_1 \ldots i_n)}(t)$ stands for the integrated variance covariance matrix whose (l, m) entry is

$$\Xi^{(i_1 \ldots i_n)}(t)_{lm} = \int_0^t \sigma_{l,i_l}(s) \sigma_{m,i_m}(s) \rho_{lm} ds$$

and ρ_{lm} is the historical correlation between S_l and S_m.

In words this joint density function mixes all of the possible volatilities and combines them as a convex combination while ensuring consistency with the initial models for the individual stocks. The same historical correlations ρ_{ij} are imposed on the densities at the constituent level but at the level of the actual process the correlations are more complex. Because of the form of the component volatilities pricing becomes immediate with this new method. Rather than employing a costly discretization scheme as was required for the conventional method, we can run a set of single step Monte Carlo integrations for each combination $(i_1, \ldots, i_n)$.

4. PRICING HIMALAYAN OPTIONS WITH MVMD

We will now apply the MVMD to the pricing of a Himalayan written on stocks that show a smile like volatility. Suppose we have a Himalayan written on two stocks S_1, S_2 over two time periods $[t_0, t_1]$ and $[t_1, t_2]$. Assuming a deterministic risk free rate, the value of the Himalayan at any time during the contract can be denoted by a generic pricing function $G(t; 0, T, A)$. The terminal payoff of the Himalayan can then be written as:

$$G(T; 0, T, A) = g(S_1(t_1), S_2(t_1), S_1(t_2), S_2(t_2)) = g(\underline{\mathbf{S}})$$

where the function $g : \mathbb{R}_+^4 \to \mathbb{R}$ represents the payout given by one of the Himalayas described earlier. The initial price is then the discounted expected value of the payout under the risk neutral measure $\mathbb{Q}$. If the joint density function of the stocks $p(S_1(t_1), S_2(t_1), S_1(t_2), S_2(t_2))$ is written as $p(\underline{\mathbf{S}})$ we get:

$$\begin{aligned} G(0; 0, T, A) &= D_{0T}\mathbb{E}[g(S_1(t_1), S_2(t_1), S_1(t_2), S_2(t_2))] \\ &= D_{0T} \int_{\mathbb{R}_+^4} g(\underline{\mathbf{S}})p(\underline{\mathbf{S}})d\underline{\mathbf{S}}. \end{aligned} \tag{9}$$

Provided that the density function $p(S_1(t_1), S_2(t_1), S_1(t_2), S_2(t_2))$ is known, this four dimensional integral can be evaluated numerically via Monte Carlo methods to give the initial price of the Himalayan.

Applying the MVMD to the Himalayan now becomes apparent. Suppose we have calibrated the volatility surfaces of our two stocks according to the mixture:

$$\begin{aligned} S_1 : \quad (\lambda_{11}, \lambda_{12}) = (0.9, 0.1), \quad (\sigma_{11}, \sigma_{12}) = (0.4, 0.1) \\ S_2 : \quad (\lambda_{21}, \lambda_{22}) = (0.8, 0.2), \quad (\sigma_{21}, \sigma_{22}) = (0.2, 0.15). \end{aligned}$$

The density function in (9) will then become:

$$\begin{aligned} &p(S_1(t_1), S_2(t_1), S_1(t_2), S_2(t_2)) = f(S_1(t_1), S_2(t_1))h(S_m(t_2)) \\ &= \lambda_{11}\lambda_{21}f^{(11)}(S_1(t_1), S_2(t_1))h(S_m(t_2)) + \lambda_{12}\lambda_{21}f^{(21)}(S_1(t_1), S_2(t_1))h(S_m(t_2)) \\ &+ \lambda_{12}\lambda_{22}f^{(22)}(S_1(t_1), S_2(t_1))h(S_m(t_2)) + \lambda_{11}\lambda_{22}f^{(12)}(S_1(t_1), S_2(t_1))h(S_m(t_2)). \end{aligned}$$

where the $f^{(ij)}$ functions are as defined in the previous section and $h(S_m(t_2))$ is the distribution of the remaining stock over the second period which will be a mixture of lognormals determined from the initial calibration. In this case the problem of incorporating the smile effect into the Himalayan now becomes one of calculating four integrals over the first period. Each of these integrals prices a Himalayan written on two underlying stocks that follow geometric Brownian motion with flat volatility smiles. The theory allows for term structure to be included in the volatility, but in our examples volatility will be kept constant. For the density $f^{(ij)}$ the auxiliary stocks will have the following dynamics over period one:

$$\begin{aligned} dS_1(t) = \mu S_1(t)dt + \sigma_{1i}S_1(t)dW_1 \\ dS_2(t) = \mu S_2(t)dt + \sigma_{2j}S_2(t)dW_2. \end{aligned}$$

The appropriate payoff will then be applied to the stocks and one of them will be eliminated. This determines the density $h(S_m)$ in each of the four integrals which is written in terms of its two constituent densities. Thus the density function will be made up of eight terms with the general component being of the form

$$\lambda_{1i}\lambda_{2j}f^{(ij)}(S_1(t_1), S_2(t_1))\lambda_{mk}h^{(k)}(S_m(t_2))$$

where over period two the auxiliary stock $S_m(t)$ follows the dynamics

$$dS_m(t) = \mu S_m(t)dt + \sigma_{mk}S_m(t)dW_m.$$

The correlation between the auxiliary stocks will be the same as the historical correlation ρ between the stocks S_1 and S_2. Accurate estimation of this parameter is crucial for pricing Himalayas with smile characteristics. In figure 1 we plot the initial price of various Himalayan options written on the stocks S_1 and S_2 over two equal time periods. In each plot the overall level of historical correlation is varied while the volatility of the individual stocks remains unchanged. These prices are calculated using one hundred thousand Monte Carlo paths. The risk free rate is fixed at 0.1 and both stocks have an initial price of 100. The plots firstly indicate that regardless of the level of the volatility smile the locally floored Himalayan of type (2) has very little correlation sensitivity. The globally floored Himalayan in (1) is long correlation for both the flat and mild volatility smile. This is an expected result since the globally floored variant is an option on the performance of the overall basket rather than an option on the individual stocks. However when the smile becomes more prominent the Himalayan shows no sensitivity to correlation. The Himalayan (6) with periodic performance is short correlation for all levels of the volatility smile. This is unsurprising since the structure is similar to a best of option on the two stocks and then an option on the single remaining stock. Both of these are short correlation so their sum will also be short correlation. In a similar manner the Himalayan variant (5) also has short correlation for both the constant volatility and smile like volatility.

It should be noted that when using the MVMD technique the actual correlation between S_1 and S_2 will not be the same as the historical correlation ρ which is used in the conventional method. The covariance of the stocks is also different. Through the multidimensional Kolmogorov equation it can be shown that in the MVMD regime the variance and covariance of our stocks S_1 and S_2 are given by:

$$C_{11}(S_1, S_2, t) = \frac{\sum_{k,k'=1}^{2} \lambda_{1k}\lambda_{2k'}\sigma_{1k}^2(t)p^{(kk')}(S_1(t), S_2(t))}{\sum_{k,k'=1}^{2} \lambda_{1k}\lambda_{2k'}p^{(kk')}(S_1(t), S_2(t))}$$

$$C_{22}(S_1, S_2, t) = \frac{\sum_{k,k'=1}^{2} \lambda_{1k}\lambda_{2k'}\sigma_{2k'}^2(t)p^{(kk')}(S_1(t), S_2(t))}{\sum_{k,k'=1}^{2} \lambda_{1k}\lambda_{2k'}p^{(kk')}(S_1(t), S_2(t))}$$

and

$$C_{12}(S_1, S_2, t) = \frac{\sum_{k,k'=1}^{2} \lambda_{1k}\lambda_{2k'}\sigma_{1k}(t)\sigma_{2k'}(t)\rho p^{(kk')}(S_1(t), S_2(t))}{\sum_{k,k'=1}^{2} \lambda_{1k}\lambda_{2k'}p^{(kk')}(S_1(t), S_2(t))}.$$

Comparing these with (7) and (8) we see that the variances of the stocks take on a new structure. The variances of the two stocks are now fully dependent on each other in a manner which is not seen in the conventional method. In fact it can be shown that the variance and covariance of the conventional method can be seen as an approximation for the MVMD technique valid for weakly correlated systems. In figure 2 we plot the conventional price and MVMD price of a type (5) Himalayan written on S_1 and S_2. For low correlations the two prices coincide but for larger correlations they diverge.

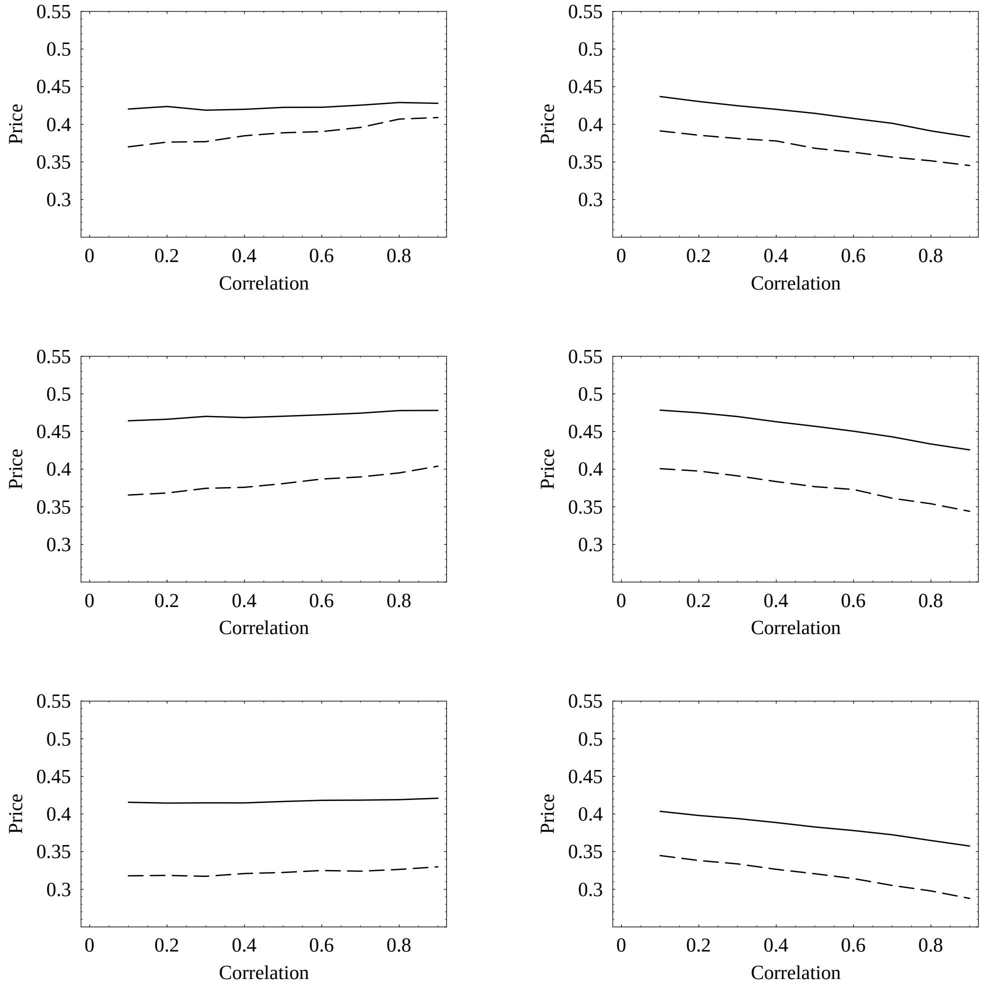

Figure 1: Price against varying correlation for Himalayans of type (1) (dashed line) and (2) in the left column and type (5) (dashed line) and (6) in the right column. Component volatilities are fixed at $(\sigma_{11}, \sigma_{12}) = (0.4, 0.1)$ and $(\sigma_{21}, \sigma_{22}) = (0.2, 0.15)$. The λ values vary. Top: $(\lambda_{11}, \lambda_{12}) = (1, 0)$ and $(\lambda_{21}, \lambda_{22}) = (1, 0)$. Middle: $(\lambda_{11}, \lambda_{12}) = (0.9, 0.1)$ and $(\lambda_{21}, \lambda_{22}) = (0.8, 0.2)$. Bottom: $(\lambda_{11}, \lambda_{12}) = (0.5, 0.5)$ and $(\lambda_{21}, \lambda_{22}) = (0.5, 0.5)$

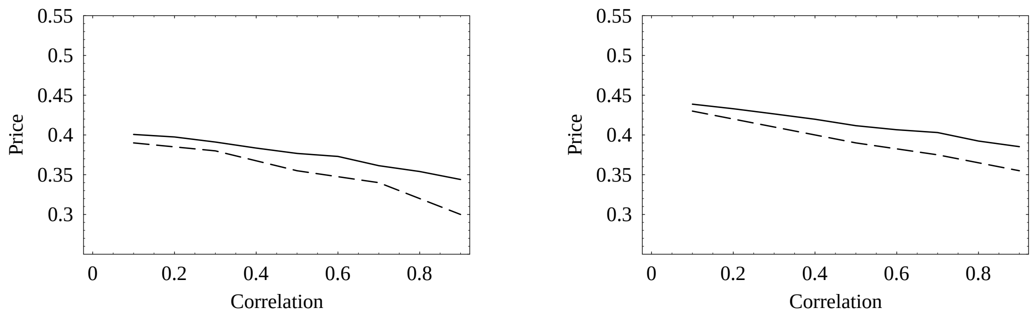

Figure 2: Price against varying correlation for the conventional method (dashed) and MVMD method. Component volatilities are fixed at $(\sigma_{11}, \sigma_{12}) = (0.4, 0.1)$ and $(\sigma_{21}, \sigma_{22}) = (0.2, 0.15)$. The λ values vary. Left: $(\lambda_{11}, \lambda_{12}) = (0.9, 0.1)$ and $(\lambda_{21}, \lambda_{22}) = (0.8, 0.2)$. Right: $(\lambda_{11}, \lambda_{12}) = (0.99, 0.01)$ and $(\lambda_{21}, \lambda_{22}) = (0.95, 0.05)$

5. CONCLUSIONS

We have outlined the MVMD technique and shown how to apply this to many variants of the Himalayan option. The main advantage of this technique is that it allows us to price options with single step Monte Carlo calculations. This is a major improvement over the the conventional mixture of densities which relies on costly time discretization and simple correlations. A comparison between the conventional MD and the MVMD has also validated theoretical results. The price of a Himalayan calculated using the two methods diverges for larger correlations. The MVMD also allows accurate modeling of stock distribution. In particular greater attention can be given to the tails which is important because of the maximization process which is embedded into the payout function. Our calculations illustrate the sensitivity of Himalayas to the historical correlation between stocks and they also show how the smile of individual stocks can change the behavior of the option. Further work could include comparing this method to stochastic volatility models which can be used to price Himalayas and studying the fit of MVMD model correlation against actual correlations.

References

P. Boyle, M. Broadie, and P. Glasserman. Monte carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21:1267–1321, 1997.

D. Brigo, F. Mercurio, and F. Rapisarda. Alternative asset-price dynamics and volatility smile. *Quantitative Finance*, 3:173–183, 2003.

D. Brigo, F. Mercurio, and F. Rapisarda. Connecting univariate smiles and baskets dynamics: a new multidimensional dynamics for basket options, 2004. URL `www.ima.umn.edu/talks/workshops/4-12-16.2004/rapisarda/MultivariateSmile.pdf`.

M. Broadie and P. Glasserman. Estimating security price derivatives by simulation. *Journal of Management Science*, 42:269–285, 1996.

O. Mahomed. Pricing of Himalayan options. Master's thesis, University of the Witwatersrand, Johnannesburg, 2004.

M. Overhaus. Himalaya options. *Risk*, 15(3):101–104, 2002.

R. Quessette. New products, new risks. *Risk*, 15(3):97–100, 2002.

T. Rådberg and O. Rundlöf. Pricing of Emerald options under local volatility. Master's thesis, Lund University, Sweden, 2006.

G. West. Exotic swaps and options, 2005. URL `www.finmod.co.za`.

AN ACTUARIAL APPROACH TO SHORT-RUN MONETARY EQUILIBRIUM

Fernando Mierzejewski

Faculty of Economy and Applied Economy, Katholieke Universiteit Leuven, Naamsestraat 69, 3000 Leuven, Belgium
Email: Fernando.Mierzejewski@wis.kuleuven.be

Abstract

The extent to which the money supply affects the aggregate cash balance demanded at a certain level of nominal income and interest rates is determined by the interest-rate-elasticity and stability of the money demand. An actuarial approach is adopted in this paper for dealing with investors facing liquidity constraints and maintaining different expectations about risks. Under such circumstances, a level of surplus exists which maximises expected value. Moreover, when the *distorted probability principle* is introduced, the *optimal* liquidity demand is expressed as a *Value at Risk* and the *comonotonic* dependence structure determines the amount of money demanded by the economy. As a consequence, the more unstable the economy, the greater the interest-rate-elasticity of the money demand. Moreover, for different parametric characterisation of risks, *market* parameters are expressed as the weighted average of *sectorial* or *individual* estimations, in such a way that multiple equilibria of the economy are possible.

1. INTRODUCTION

According to the Keynes's *liquidity preference* proposition, the demand for cash balances is positively affected by the level of income and negatively affected by the return offered by a class of money substitutes, see Keynes (1935). The first part of the proposition is a consequence of the assumption that the amount of transactions is proportional to the level of income. To explain the effect of the interest rate, Keynes emphasises the influence of capital fluctuations in decision-making. Thus, investors expecting interest rates to rise demand fewer risk-free securities in order to avoid capital losses — since the price of such instruments is expected to diminish in this case. By contrast, when interest rates are expected to fall, more bonds are demanded — in this way, capital gains can be attained after the collapse of interest rates. Therefore, fewer provisions are maintained for high levels of the interest rate and vice versa.

In macroeconomic analysis, the level of prices establishes the connection between *nominal* magnitudes, expressed in monetary units, and *real* quantities, which represent flows of goods and services. Accordingly, $Y = P \cdot y$, where P, Y and y respectively denote the level of prices, the nominal income and the real income. Let us additionally denote by M the total money supply. Therefore, the short-run monetary equilibrium is given by the *quantity equation*:

$$M = P\, y \cdot l(r) = Y \cdot l(r). \tag{1}$$

The *liquidity preference* function $l(r)$ expresses the ratio between demanded cash balances and nominal income. It is not likely to be constant but it may change slowly over time. The inverse ratio of the liquidity preference function is called *velocity of money*.

Any change in the money supply will require a change in one or more of the variables determining the liquidity demand (i.e. P, y or r) in order to reestablish the monetary equilibrium. When prices are rigid for short-run fluctuations and the real product remain stable in short-terms, the whole adjustment is performed in $l(r)$. In addition, if the liquidity preference is *absolute*, i.e. if investors are satisfied at a single level of the interest rate,[1] the amount of money can change without a change in either nominal income or interest rates. Under such circumstances, monetary policy is useless for dealing with short-run fluctuations. The situation is different if prices are flexible and liquidity preference is *non-absolute*. Then a monetary expansion produces a new equilibrium involving a higher price for the same quantity, the higher this response the more inelastic the money demand. In the short-run, production is encouraged until prices are reestablished at their original level. In the long-run, new producers enter the market and existing plants are expanded, as claimed by Friedman (1970).

Under such circumstances, the efficacy of monetary policy depends on the degree of rigidity of prices and the elasticity of the money demand, as well as on the stability of liquidity preference. There is a consensus among researchers about the existence of a stable long-run relationship, though fluctuations of cash balances in the short-run remain unexplained. Episodes like the *missing money* in the mid-seventies, the great velocity decline in the early eighties, followed by the expansion of narrow money in the mid-eighties, or the *velocity puzzle* of the mid-nineties, still lack a satisfactory explanation, see Ball (2001) and Ball (2002), Carpenter and Lange (2002) and Teles and Zhou (2005). In accounting for such drawbacks, recent literature has focused on *uncertainty*, which is supposed to have been incremented after 1980 due to deregulation and financial innovations, as in Atta-Mensah (2004), Baum et al. (2002), Carpenter and Lange (2002), Choi and Oh (2003) and Greiber and Lemke (2005). Deregulation and financial innovation are also given as arguments to support the role of the opportunity cost in accounting for unexplained fluctuations, see Ball (2002), Collins and Anderson (1998), Duca (2000), Dreger and Wolters (2006) and Teles and Zhou (2005). According to this view, a stable long-run relationship exists and movements of the interest rate can explain all short-run episodes, as long as the right monetary aggregate is used.

In this paper, an extended model is proposed according to which liquidity preference is explicitly determined by *uncertainty* and *information*. First, the cash demand of a single representative investor is obtained. Investors are supposed to face liquidity constraints and consequently, in *Section 2* equity is treated as an additional liability. In addition, the behaviour towards risk is determined by the transformation of probabilities according to an *informational* parameter. Then

[1] *Absolute* liquidity preference corresponds to the case when the liquidity demand is perfectly elastic with respect to the interest rate. According to Keynes, the degree of elasticity depends on how homogeneous expectations are, where perfect elasticity is obtained when expected and actual values are the same. In this case, money and risk-free securities are perfect substitutes — since no capital gains or losses are expected.

the expected return of the fund is maximised when the mathematical expectation of the residual exposure (a measure of the cost of assuming bankruptcy) plus the opportunity cost of capital is minimised. In this way, I follow Dhaene et al. (2003) and Goovaerts et al. (2005), who on these terms develop a mechanism for capital allocation. When looking for the aggregated surplus in *Section 3*, capital is supposed to be provided by a central authority or financial intermediaries acting in a competitive market, in such a way that a single interest rate is required for lending. Hence the situation is similar to the case of a centralised conglomerate distributing capital among subsidiaries, as in Dhaene et al. (2003), Goovaerts et al. (2005) and Mierzejewski (2006), and the opportunity cost of money is related to the average return over a class of money substitutes. Thus, monetary aggregates are determinants of liquidity preference in the model. Finally, within a Gaussian setting, the aggregate exposure is normally distributed and its volatility is equal to the weighted average of individual volatilities. Therefore, aggregation plays a role in the determination and stability of the liquidity demand. The same results are obtained when marginal risks are exponentially and Pareto distributed. The final remarks are given in *Section 4*.

2. THE RATIONAL MONEY DEMAND

Since in frictionless markets the amount of cash maintained for precautionary purposes can be modified at any time by lending and borrowing, managers who maximise value demand no equity — which is actually the proposition established by Modigliani and Miller (1958). However, averse-to-risk customers are sensible to fluctuations and, as long as the business activities of financial intermediaries — which accordingly are said to be *opaque* — are not observed by outsiders, a pressure is established to be perceived as default-free, as suggested by Merton (1997). In the model developed by Tobin (1956), averse-to-risk investors show liquidity preference as behaviour towards uncertainty. Assuming that risks follow Gaussian distributions, a linear relationship is established between the expected returns and volatilities of the portfolios containing a proportion of a certain fund and a cash guarantee, which determines the set of *efficient* portfolios — in the sense that for any combination outside the line, it is always possible to build a new fund providing the same expected return and a lower risk, or the same risk but a higher return. The way preferences affect portfolio decisions can then be analysed in the plane of expected returns and volatilities, where the indifference curves of risk-lovers should present a negative slope, as long as such individuals accept a lower expected return if there is a chance to obtain additional gains. By contrast, averse-to-risk investors do not take more risk unless they are compensated by a greater expected return and consequently, their indifference curves have positive slopes. Therefore, for any risk-aversion profile, the optimal combination is determined by the (tangency point of) intersection between the unique indifference curve representing preferences and the line of efficient portfolios.

Let us analyse in the following how the Tobin's model is affected by the hypothesis of *imperfect* competition, a case where risks belong to a general class of probability distributions which economic agents distort according to their information and knowledge when making decisions. Moreover, liquidity constraints are faced when borrowing and lending and managers have to expend effort to correctly assess prices. Let the parameter θ denote the state of information of an investor holding a mutual fund whose percentage return is represented by the random variable X.

Because of the precautionary motive, a guarantee L is maintained for a determined period of time to avoid bankruptcy. In order to introduce in the model the effect of liquidity constraints, equity is regarded as an additional liability and the size of the guarantee is expressed as a proportion of the level of income $\overline{Y}$, such that $L = \overline{Y} \cdot l$, where l represents the proportion of income assigned to the non-risky asset. Hence, if r_0 denotes the risk-free interest rate, the percentage capital return of the total portfolio can be expressed as $Y = X - l - r_0 \cdot l$ and decisions are affected by the *percentage* return on income:

$$\mu_{\theta,Y} = E_\theta[Y] = (\mu_{\theta,X} - l) - r_0 \cdot l.$$

In giving a meaning to the informational parameter θ, let us stress the fact that *expectations* are wanted to be modified. Then probability beliefs are transformed by a distortion parameter which is supposed to be determined by *information* and *knowledge* and the *proportional hazards distortion* is introduced, see Wang (1995):

$$\mathrm{E}_\theta \left[X\right] = \int x \; dF_{\theta,X}(x) = \int G_{\theta,X}(x) \; dx := \int G_X(x)^{\frac{1}{\theta}} \; dx.$$

The *cumulative* and *decumulative* (also known as *survival*) *probability distribution* functions have been introduced, $F_{\theta,X}(x) = P_\theta \left[X \le x\right] = 1 - P_\theta \left[X > x\right] = 1 - G_{\theta,X}(x)$. When $\theta > 1$, the expected value of risk is overestimated and it is underestimated when $\theta < 1$, in this way respectively accounting for the behaviour of *averse-to-risk* and *risk-lover* investors.

Notice, however, that individuals react differently depending on the sign of the capital return. In fact, when a loss is suffered, cash is demanded to avoid default, while in the case a gain is obtained the surplus can be used to pay current liabilities or assigned to new investments. Hence, decision-makers mainly concerned about the speculative and the precautionary motives respectively focus on the terms $E_\theta \left[(X - l)_+\right]$ and $E_\theta \left[(X + l)_-\right]$. Let us accordingly assume that capital decisions are taken by risk managers who minimise bankruptcy and rely on the *average* value of the insured return:

$$E_\theta \left[\, (X - l)_+\right] \;\approx\; E_\theta \left[X_+\right] - r_{\theta,X} \cdot l.$$

Since the term $r_{\theta,X} > 0$ represents the absolute value of the marginal reduction in insured capital gains produced when attracting an additional unit of equity, it can be regarded as a *premium for solvency*. Hence the following expression is obtained for the expected percentage income:

$$\mu_{\theta,Y} = E_\theta \left[X_+\right] - E_\theta \left[\, (X + l)_-\right] - (r_0 + r_{\theta,X}) \cdot l.$$

Under such conditions, *precautionary* investors that maximise value minimise bankruptcy costs. Applying Lagrange optimisation, we obtain that decision makers attract funds until the marginal return of risk equals the total cost of capital:

$$-\frac{\partial}{\partial l} \, E_\theta \left[\, (X + l)_-\right] - (r_0 + r_{\theta,X}) = G_{\theta,-X} \left(l^*\right) - (r_0 + r_{\theta,X}) = 0.$$

Equivalently, it can be said that investors stop demanding money at the level at which the marginal expected gain in solvency equals its opportunity cost. Thus the optimal cash demand is given by:

$$l_{\theta,X} \left(r_0 + r_{\theta,X}\right) \;=\; G^{-1}_{\theta,-X} \left(r_0 + r_{\theta,X}\right). \qquad (2)$$

From this expression, the money demand follows a decreasing and — as long as the distribution function describing uncertainty is continuous — continuous path, whatever the kind of risks and distortions. The minimum and maximum levels of surplus are respectively demanded when $(r_0 + r_{\theta,X}) \geq 1$ and $(r_0 + r_{\theta,X}) \leq 0$.

In practical applications, intermediaries face operational and administrative costs, at the time that a premium over the risk-free interest rate is asked for lending in secondary markets. Hence, the return $r_0 + r_{\theta,X}$ can be interpreted as a *net* opportunity cost. Though environmental facts, such as the perception of credit quality and gains in efficiency because of improvements on analysis and administration, are expected to evolute on time, we can regard them as softly modified — and not a matter of *speculation*. Also the risk attitude of managers is supposed to remain more or less unchanged. Therefore, the parameter θ is expected to remain stable and consequently, as long as the probability distribution of the random variable X is also stable, the capital decisions of investors should remain more or less the same and the economy as a whole should behave accordingly.

However, if probability distributions are allowed to evolve on time — i.e. if the processes of capital gains and losses are not *stationary* — so does the premium for solvency $r_{\theta,X}$. Actually, this can be the case after a monetary expansion — which can be performed by the central bank as well as by the entrance of new investors — since as long as part of the extra money is used to buy financial securities and the increment in demand is high and persistent enough to induce the price to rise *more frequently,* the term $E_\theta\left[(X-l)_+\right]$ is pushed to increase. In a similar way, a monetary contraction can press the insured return to decrease. This situation might in turn impel decision-makers to actualise expectations and so the informational parameter θ might be modified. But this adjustment is supposed to be produced with a certain delay — for time is required for analysis — while the opportunity cost may be *instantaneously* altered. Therefore, changes in the stock of money may induce instability *from within* in secondary markets. Adjustments are performed along a stable money demand relationship, though the process may be reinforced by structural modifications once expectations are actualised.

3. SHORT-RUN MONETARY EQUILIBRIUM

In order to obtain an expression for the cash balance demanded by the whole economy, let us assume that economic agents hold aggregate exposures characterised by the random variables $X_1, \ldots, X_n$. Capital is supplied by a central authority at a single interest rate r (or, equivalently, secondary markets are regarded as competitive and financial intermediaries are *price takers*) relying on the informational parameter θ and the uncertainty introduced by the *market* portfolio X. When different expectations are allowed among decision makers, the aggregate money demand is given by:

$$l_{\theta_1,\ldots,\theta_n,-X}(r) \;=\; \sum_{i=1}^{n} G^{-1}_{\theta_i,-X_i}(r) \;=\; G^{-1}_{\theta_1,\ldots,\theta_n,-X}(r).$$

The second equality is a mathematical identity as long as the process of capital gains and losses of the market portfolio is described by the *comonotonic sum* $X = X_1^c + \cdots + X_n^c$, where $G_{\theta_1,\ldots,\,\theta_n,-X} = \left(\sum_{i=1}^{n} G^{-1}_{\theta_i,-X_i}\right)^{-1}$ denotes the distribution function of the comonotonic sum when marginal distributions are given by $(G_{\theta_1,-X_1}, \ldots, G_{\theta_n,-X_n})$. *Comonotonicity* characterises an extreme case of dependence, when no benefit can be obtained from diversification.[2] Thus *precautionary* investors rely on the most pessimistic case, when the failure in any single firm spreads all over the market.

The dependence of the liquidity demand on the variability of income becomes explicit in a Gaussian setting. Let us assume in the following that individual exposures are distributed as Gaussians with means $\mu_1, \ldots, \mu_n$ and volatilities $\sigma_1, \ldots, \sigma_n$, while the contributions of individual exposures to the market portfolio are given by the coefficients $\lambda_1, \ldots, \lambda_n$, with $0 \leq \lambda_i \leq 1 \; \forall i$, such that $\overline{Y}_i = \lambda_i \cdot \overline{Y}$ and $\overline{Y} = \overline{Y}_1 + \cdots + \overline{Y}_n$. Volatilities are expressed as proportions of the levels of income and can be interpreted as the volatilities of different funds as well as the distorted volatilities of the same Gaussian exposure — or some intermediate case. Under such conditions, the comonotonic sum is also a Gaussian random variable, see Dhaene et al. (2002), whose mean and volatility are respectively given by:

$$\mu \;=\; \sum_{i=1}^{n} \lambda_i \cdot \mu_i \qquad \& \qquad \sigma \;=\; \sum_{i=1}^{n} \lambda_i \cdot \sigma_i. \tag{3}$$

On these grounds, the weighted average mean and volatility describe the uncertainty of the market portfolio. In particular, high volatility may be induced by a single group, as a negative externality to more efficient companies and so the possibility of *contagion* naturally arises in the model. In the same way, stability may be inherited by less efficient institutions when low volatility predominates.

Since the *quantile* function of a Gaussian random variable can be expressed in terms of the standard Normal distribution Φ (see Dhaene et al. (2002)), the short-run monetary equilibrium is described by the following equation:

$$M \;=\; \overline{Y} \cdot l_{\mu,\sigma}(r) \;=\; \overline{Y} \cdot \Big[-\big(\mu + \sigma\, \Phi^{-1}(r)\big) \Big]. \tag{4}$$

Therefore, the monetary equilibrium can be reestablished by modifying the level of nominal income $\overline{Y}$, the average return μ, the market volatility σ or the interest rate r. As already stated, only r is expected to change in the short-run. Monitoring and analysis induce investors to eventually incorporate the new regime of X in decision making and possibly modify expectations, both determinants of μ and σ.

The difference between the classic and the extended model can be noticed by comparing *Equations 1* and *4*. Thus, while in *Equation 1* the elasticity of income with respect to the stock of money exclusively depends on the interest rate through the liquidity preference function, in *Equation 4* it is also affected by uncertainty. In addition, if $\overline{y}$ represents the level of *real* income, the *new* short-run equilibrium can be written in real terms as:

$$M \;=\; P\,\overline{y} \cdot \Big[-\big(\mu + \sigma\, \Phi^{-1}(r)\big) \Big].$$

Therefore, to stabilise the product it is also required to control the market risk. A proper monetary policy should then consider a combination of P, μ, σ and r compatible with a given level of income. The level of σ that preserves the monetary equilibrium for given values of M, $\overline{Y}$, μ and r can be regarded as the *induced* volatility. A tentative criterion for monetary policy may then involve the determination of the level of interest rates ensuring a given inflation and induced market

[2] The inverse probability distribution of the comonotonic sum is given by the sum of the inverse marginal distributions, see Dhaene et al. (2002).

volatility. Additionally, the *non-distorted* volatility can be estimated by the *standard deviation* of the random variable X representing the capital losses of the market portfolio. A measure of the degree of distortion performed by the market is thus determined by the difference between the induced and the non-distorted volatility.

An alternative representation is obtained by considering that individual exposures are *exponentially distributed*. In this case, the comonotonic sum is also exponentially distributed (see Dhaene et al. (2002)), such that if $\beta_1, \ldots, \beta_n$ denote the *informational types* of investors, with $\beta_i \geq 0 \, \forall i$, and, as before, $\lambda_1, \ldots, \lambda_n$ represent the marginal contributions to the aggregate income, with $0 \leq \lambda_i \leq 1 \, \forall i$, then the exponential parameter of the market portfolio is expressed as the weighted average of the exponential parameters of marginal risks:

$$\beta \;=\; \sum_{i=1}^{n} \lambda_i \cdot \beta_i .$$

Therefore, the liquidity preference function of the economy is given by:

$$l_\beta(r) = -\beta \cdot \ln(r), \quad \text{with} \;\; \beta > 0. \tag{5}$$

Thus, the higher the parameter β, which within this framework completely characterises risk, the more sensitive is liquidity preference to the cost of capital. In this way, uncertainty is explicitly related to the monetary equilibrium and hence to the terms of liquidity — determined by the money supply.

When marginal risks are *Pareto distributed*, the survival probability distributions as estimated by decision-makers are given by:

$$G_{-X_i}(x) = x^{-\frac{1}{\alpha_i}}, \quad \text{with} \;\; \alpha_i > 0 \;\; \text{and} \;\; x > 1.$$

The parameters $\alpha_1, \ldots, \alpha_n$ correspond to the *states of information* of investors. As long as they agree on a single value α, the comonotonic sum is also Pareto distributed (see Dhaene et al. (2002)) and liquidity preference is given by:

$$l_\alpha(r) = n \cdot r^{-\alpha}. \tag{6}$$

Under different expectations, the comonotonic sum is not necessarily Pareto distributed. However, an estimation of the parameter α can be found such that *Equation 6* determines the monetary equilibrium. In this case, the point interest-rate-elasticity is constant and equal to α. Many models for the estimation of the money demand are supported on this assumption.

4. CONCLUSIONS

An *extended* model is presented in this paper — also referred to as the *imperfect competition* model — to characterise the liquidity preference of investors facing liquidity constraints. Under such circumstances, a level of surplus exists that maximises value and the *rational* money demand is determined by the *quantile* function — a measure of the probability accumulated in the *tail* of the distribution function — of the random variable representing the series of capital profits and losses of the residual exposure (*Equation 2*). In this way, an equivalence is established between a confidence level and the opportunity cost of capital and the optimal amount of cash is determined by the exchange of a sure return and a flow of probability. An informational parameter, affecting the opportunity cost of money, represents the expectations of decision makers. Averse-to-risk and risk-lover investors respectively under and overestimate the cost of capital and so they respectively demand more and less equity.

The importance attached to liquidity preference in macroeconomic analysis is a consequence of the fact that it determines the short-run monetary equilibrium of the economy. In the classic approach, the amount of money which is compatible with given levels of nominal income and interest rates can be obtained from *Equation 1*, see Friedman (1970). According to the *extended* model presented in this paper, the aggregate money demand of the economy is given by the sum of the liquidity preferences of investors, mathematically characterised by the *comonotonic sum* of individual exposures. The aggregate money demand is thus expressed as a Value-at-Risk but referred to a *market portfolio* which relies on the most pessimistic case, when no gain can be obtained from diversification. In a Gaussian setting, the comonotonic sum is also a Gaussian variable, whose volatility is equal to the weighted average of individual volatilities (*Equation 4*). In this way, the classic model is extended allowing a correction for risk.

Within the *imperfect competition* framework, the total stock of money M, the level of income $\overline{Y}$, the interest rate r, the mean μ and the market volatility σ are all determinants of the short-run monetary equilibrium (*Equation 4*). Thus, as long as part of the funds available in the economy are spent on capital assets, an adjustment in the opportunity cost r is expected in the short-run — stimulated by the modification of the stochastic nature of capital gains — which is supposed to *instantaneously* affect liquidity preference. In the medium-term, investors correct their expectations and so part of the adjustment may be performed through *informational* shocks affecting the aggregate mean or the market volatility. An important feature of the mechanism is that the evolution of risks, motivated by flows of funds, determines expectations and not the opposite, though liquidity preference might also be affected by a *purely* informational shock.

As pointed out in *Section 2*, liquidity preference is not affected in the same way by capital gains and losses. Thus, while positive returns affect the opportunity cost of money and so determine a movement along a stable relationship, the precautionary attitude of decision makers depends on negative returns, as does the *shape* of the money demand (see *Equation 2*). The first adjustment is supposed to occur instantaneously, while the second one is performed gradually, for it takes time for investors to internalise new market conditions. In practice, both decisions are related to different markets. Accordingly, the cost of equity r is represented by the average return over a class of securities, other than cash, that can be regarded as substitutes to money. On the other hand, the liquidity preference function depends on the series of returns over a set of instruments that are representative for the assumed exposures. Then the variability showed by a representative index of this class determines the *market volatility* σ.

Finally, as stated in *Equation 3*, in a Gaussian setting the expected value of the market portfolio and the market variability are respectively given by the weighted average of individual means and volatilities. Hence, the market uncertainty will be mainly determined by a single institution or sector, in the case it contributes more to the aggregate exposure. Stability can be induced in the whole market in this way. The same results are obtained when individual exposures are *exponentially* or *Pareto* distributed, for in both cases the risk parameters are aggregated when accounting for market behaviour. Moreover, the model accepts multiple equilibria, since different combinations of the risk parameters may lead to the same market characterisation.

The terms under which market shocks affect individual expectations about risks will be determined by specific conditions, such as the state of aggregation, restrictions in the access to credit, the distribution of information within the market and the skills and knowledge of investors. Thus, changes in the aggregate monetary stock may induce intermediaries to prefer bigger or more efficient companies — *flight to quality* — a situation that may become more difficult according to the availability of funds and possibly increment more the riskiness of less productive sectors of the economy. In this way, within the *imperfect competition* framework, a broader meaning is attached to *instability*.

References

J. Atta-Mensah. Money demand and economic uncertainty. Technical Report 04-25, Bank of Canada, July 2004.

L. Ball. Another look at long-run money demand. *Journal of Monetary Economics*, 47(1):31–44, 2001.

L. Ball. Short-run money demand. Technical Report 481, The Johns Hopkins University, Department of Economics, August 2002.

C.F. Baum, M. Caglayan, N. Ozkan, and O. Talavera. The impact of macroeconomic uncertainty on non-financial firms demand for liquidity. Technical Report 552, Boston College Department of Economics, Dec 2002.

S.B. Carpenter and J. Lange. Money demand and equity markets. Technical Report No. 2003-03, Board of Governors of the Federal Reserve System, October 2002.

W.G. Choi and S. Oh. A money demand function with output uncertainty, monetary uncertainty, and financial innovations. *Journal of Money, Credit and Banking*, 35:685–709, 2003.

S. Collins and R. Anderson. Modeling U.S. households' demands for liquid wealth in an era of financial change. *Journal of Money, Credit and Banking*, 30(1):83–101, 1998.

J. Dhaene, M. Denuit, M. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics & Economics*, 31:3–33, 2002.

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

C. Dreger and J. Wolters. Investigating the M3 money demand in the euro area — new evidence based on standard models. Technical Report 561, German Institute for Economic Research, March 2006.

J.V. Duca. Financial technology shocks and the case of missing M2. *Journal of Money, Credit and Banking*, 32(4):820–839, 2000.

M. Friedman. A theoretical framework for monetary analysis. *The Journal of Political Economy*, 78(2):193–238, 1970.

M.J. Goovaerts, E. Van den Borre, and R. Laeven. Managing economic and virtual economic capital within financial conglomerates. *North American Actuarial Journal*, 9(3):77–89, 2005.

C. Greiber and W. Lemke. Money demand and macroeconomic uncertainty. Technical Report 26, Deutsche Bundesbank, Research Centre, August 2005.

J.M. Keynes. *The General Theory of Employment, Interest and Money.* Cambridge University Press, 1935.

R.C. Merton. A model of contract guarantees for credit-sensitive, opaque financial intermediaries. *European Finance Review*, 1(1):1–13, 1997.

F. Mierzejewski. Economic capital allocation under liquidity constraints. In M. Vanmaele, A. De Schepper, J. Dhaene, H. Reynaerts, W. Schoutens, and P. Van Goethem, editors, *Proceedings of the 4th Actuarial and Financial Mathematics Day*, pages 107–116, Brussels, 2006. Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten.

F. Modigliani and M.H. Miller. The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, 48:261–297, June 1958.

P. Teles and R. Zhou. A stable money demand: Looking for the right monetary aggregate. *Economic Perspectives*, 1Q/2005:50–63, 2005.

J. Tobin. Liquidity preference as behavior toward risk. Technical Report 14, Cowles Foundation, Yale University, July 1956. Discussion Paper.

S. Wang. Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics & Economics*, 17(1):43–54, 1995.

AVERAGED BOND PRICES IN GENERALIZED COX-INGERSOLL-ROSS MODEL OF INTEREST RATES

Beáta Stehlíková[†1]

[†]*Department of Applied Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynska dolina, Bratislava, Slovakia*
Email: `stehlikova@fmph.uniba.sk`

Abstract

In short rate interest models, the behaviour of the short rate is given by a stochastic differential equation (in one-factor models) or a system of stochastic differential equations (in multi-factor models). Interest rates with different maturities are determined by bond prices, which are solutions of the parabolic partial differential equation. We consider the generalized Cox-Ingersoll-Ross model, where the short rate is a sum of two Bessel square root processes, which evolve independently. The bond price is a function of maturity and of the level of each of the components of the short rate. We do not observe all values necessary to obtain a bond price. Therefore, we propose the averaging of the bond prices. We consider the limiting distribution of the short rate components, conditioned to have the sum equal to the observable short rate level. In this way, we obtain the averaged bond prices, which depend only on maturity and short rate. We prove that there is no one-factor model yielding the same bond prices as with the averaged values described above.

1. GENERALIZED COX-INGERSOLL-ROSS MODEL OF INTEREST RATES

Term structure models describe the dependence between the time to maturity of a discount bond and its present price which implies the interest rate. Continuous short rate models are formulated in terms of one or more stochastic differential equations for the instanteneous interest rate r (short rate). The bond prices, and hence the term structures of the interest rates, are then obtained by solving the partial differential equation.

In one-factor models, the process describing the short rate, is given by

$$dr = \alpha(t,r)dt + \beta(t,r)dw, \tag{1}$$

[1]The author was supported by the VEGA grant 1/3767/06.

where $\alpha(t,r)$ and $\beta(t,r)$ are non-stochastic functions, and w is a Wiener process. If $\alpha(t,r) = \kappa(\theta - r)$, $\kappa > 0$, the process has the property of mean-reversion to the level θ. A popular class of models is obtained by taking $\sigma(t,r) = \sigma r^{\gamma}$. It includes the Vasicek model [$\gamma = 0$, see Vasicek (1977)], and the Cox-Ingersoll-Ross (CIR) model [$\gamma = \frac{1}{2}$ see Cox et al. (1985)]; for a comparison of models with different values for γ, see e.g. Chan et al. (1992).

If the short rate evolves according to 1, then the discount bond with maturity T has the price $P(t,r)$ depending on the time t and the current level of the short rate r. It is given by the following partial differential equation:

$$-\frac{\partial P}{\partial t} + (\alpha - \lambda\beta)\frac{\partial P}{\partial r} + \frac{1}{2}\beta^2\frac{\partial^2 P}{\partial r^2} - rP = 0, \quad t \in (0,T) \tag{2}$$

$$P(T,r) = 1, \tag{3}$$

where $\lambda = \lambda(t,r)$ is the market price of risk. The interest rates are then obtained from the bond prices by $R(t,r) = -\frac{\log P(t,r)}{T-t}$, see Kwok (1998).

For the specific choices of the market price of risk in Vasicek and CIR models, it is known that the bond price can be written in closed form. If $\lambda(t,r) = \lambda\sqrt{r}$ in the CIR model then the price of the bond with time to maturity $\tau = T - t$ has the form

$$P(\tau,r) = A(\tau)e^{-B(\tau)r}.$$

The functions $A(\tau)$ and $B(\tau)$ satisfy the following system of ordinary differential equations

$$\begin{aligned} \dot{A}(\tau) &= \kappa\theta A(\tau)B(\tau) \\ \dot{B}(\tau) &= -(\kappa + \lambda\sigma)B - \frac{1}{2}\sigma^2 B(\tau)^2 + 1 \end{aligned} \tag{4}$$

with initial conditions $A(0) = 1$, $B(0) = 0$, which can be solved analytically.

There are several possibilities of generalizing one-factor models, leading to multifactor models: making a parameter of the one-factor model stochastic (e.g. stochastic volatility models Anderson and Lund (1996), Fong and Vasicek (1991)), adding extra relevant quantities (consol rate in Brennan and Schwartz (1982), European interest rate in Corzo Santamaria and Schwartz (2000), Corzo Santamaria and Biscarri (2005)), composing the short rate by means of more components (generalized CIR model in Cox et al. (1985), consol rate and the spread between the short rate and consol rate in Schaefer and Schwartz (1984), Christiansen (2002)), etc.

In the generalized CIR model, the short rate r is the sum of two independent Bessel square root processes:

$$\begin{aligned} r &= r_1 + r_2, \\ dr_1 &= \kappa_1(\theta_1 - r_1)dt + \sigma_1\sqrt{r_1}dw_1, \\ dr_2 &= \kappa_2(\theta_2 - r_2)dt + \sigma_2\sqrt{r_2}dw_2, \end{aligned} \tag{5}$$

where the Wiener processes w_1 and w_2 are independent. If the market prices of risk corresponding to r_1 and r_2 are taken to be $\lambda_1\sqrt{r_1}$ and $\lambda_2\sqrt{r_2}$, then the bond price $\pi(\tau, r_1, r_2)$ has the form

$$\pi(\tau, r_1, r_2) = A(\tau)e^{-B_1(\tau)r_1 - B_2(\tau)r_2}, \tag{6}$$

where $A(\tau) = A_1(\tau)A_2(\tau)$ with $A_1(\tau)$, $A_2(\tau)$, $B_1(\tau)$, $B_2(\tau)$ the solutions of the systems of ordinary differential equations (4) arising in the one-factor model, with the appropriate index.

2. AVERAGING IN TWO-FACTOR MODELS

Since the components of the short rate r_1 and r_2 are not observable and the observable variable is only their sum r, an interesting question refers to the properties of the average of the two bond prices conditioned on the given sum of r_1 and r_2. This is motivated by several papers: e.g. Fouque et al. (2003) with an averaging in stochastic volatility models of stock prices, or Cotton et al. (2004) with an averaging in stochastic volatility models of bond prices (where the unobservable random quantity is the volatility), which are used in the series expansion of the prices. The asymptotic distribution of the hidden process is used. It can be justified if the processes have been evolving for a sufficiently long time.

In the same way, we consider the limit distributions in the generalized CIR model. It is well known that the limit distribution of a Bessel square root process is a gamma distribution. Hence the limit distribution of each of the two rates r_i ($i = 1, 2$) in (5) is given by

$$f_i(r_i) = \frac{a_i^{b_i}}{\Gamma(b_i)} e^{-a_i r_i} r_i^{b_i - 1}$$

where $a_i = \frac{2\kappa_i}{\sigma_i^2}$, $b_i = \frac{2\kappa_i\theta_i}{\sigma_i^2}$ for $r_i > 0$ and zero otherwise. The limit density of r_1 conditioned on $r_1 + r_2 = r$ is equal to

$$f(r_1, r) = \frac{f_1(r_1)f_2(r - r_1)}{\int_0^r f_1(s)f_2(r - s)ds} = \frac{f_1(r_1)f_2(r - r_1)}{M(r)}, \tag{7}$$

where we used $M(r)$ for the denominator of the fraction in order to simplify the notations in the computations hereafter. The bond price (6) can be written in terms of τ, r, r_1 and the averaged value is computed as

$$P(\tau, r) = \int_0^r \pi(\tau, r_1, r - r_1) f(r_1, r) dr_1. \tag{8}$$

In the same way, the averaged term structure is given by

$$P(\tau, r) = \int_0^r \left[-\frac{\log \pi(\tau, r_1, r - r_1)}{\tau} \right] f(r_1, r) dr_1. \tag{9}$$

In Fig. 1 we give an example, by showing the term structures obtained by the generalized CIR model and the averaged term structure computed as described above.

3. THE MAIN RESULT

In this paper, we study the following problem: is it possible to find functions α and β, see equation (1), such that the bond prices are the same as the averaged prices from a two-factor CIR model as in (5).

We restrict ourselves to a specific class of processes:

- drift and volatility of the process, as well as the market price of risk are time-independent;

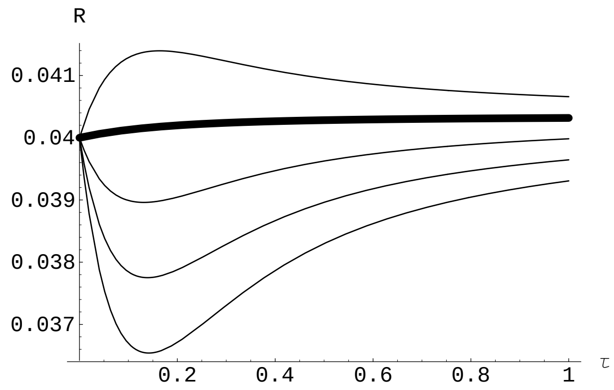

Figure 1: Examples of term structures corresponding to different pairs of r_1 and r_2 such that $r_1 + r_2 = 0.04$. The averaged term structure is in bold.

- for a zero level of short rate, we require the volatility to be zero; this condition ensures the nonnegativity of the short rate;
- the volatility parameters σ are different for the two processes forming the short rate in the two-factor CIR model.

We start with the following result.

Theorem 3.1 *If we assume that the functions α, β, and λ only depend on r (and not on τ), that the functions α, β, and λ are continuous in $r = 0$, that $\beta(0) = 0$, and that $\sigma_1 \neq \sigma_2$, then*

(a) $P(\tau, r) \to A(\tau)$ *as* $r \to 0$,

(b) $\frac{\partial P}{\partial \tau}(\tau, r) \to \dot{A}(\tau)$ *as* $r \to 0$,

(c) $\frac{\partial P}{\partial r}(\tau, r) \to -A(\tau)\left(\frac{b_1}{b_1+b_2} B_1(\tau) + \frac{b_2}{b_1+b_2} B_2(\tau)\right)$ *as* $r \to 0$,

(d) $\frac{\partial^2 P}{\partial r^2}(\tau, r)$ *is bounded in the neighbourhood of* $r = 0$.

For the proof, we need some properties of the Kummer confluent hypergeometric functions ${}_1F_1$, which are recalled in the following lemma, see Abramovitz and Stegun (1972).

Lemma 3.2 *The following equalities hold:*

- $\displaystyle\int_0^r e^{-ax} x^{b-1} (r-x)^c dx = r^{b+c} \frac{\Gamma(b)\Gamma(1+c)}{\Gamma(1+b+c)} \, {}_1F_1(b, 1+b+c, -ar)$
- $\displaystyle {}_1F_1(a, b, z) = 1 + \frac{a}{b} z + \frac{a(a+1)}{b(b+1)} z^2 + \cdots .$

Proof of theorem 3.1:
Firstly, we rewrite both $M(r)$ and the density $f(r_1, r)$ by means of confluent hypergeometric functions:

$$M(r) = \int_0^r f_1(r_1) f_2(r - r_1) dr_1$$
$$= \frac{a_1^{b_1} a_2^{b_2}}{\Gamma(b_1 + b_2)} e^{-a_2 r} r^{b_1 + b_2 - 1} {}_1F_1(b_1, b_1 + b_2, -(a_1 - a_2) r)$$

and

$$f(r_1, r) = \frac{1}{M(r)} f_1(r_1) f_2(r - r_1) \tag{10}$$
$$= \frac{1}{{}_1F_1(b_1, b_1 + b_2, -(a_1 - a_2) r)} \frac{\Gamma(b_1 + b_2)}{\Gamma(b_1)\Gamma(b_2)} \frac{1}{r^{b_1 + b_2 - 1}} \left[e^{-(a_1 - a_2) r_1} r_1^{b_1 - 1} (r - r_1)^{b_2 - 1} \right].$$

Now, we can prove the assertions of the theorem.

(a) Substituting (10) into the expression for the averaged bond price gives

$$P(\tau, r) = \int_0^r \pi(\tau, r_1, r - r_1) f(r_1, r) dr_1$$
$$= A e^{-Br} \frac{{}_1F_1(b_1, b_1 + b_2, -((B_1 - B_2) + (a_1 - a_2) r))}{{}_1F_1(b_1, b_1 + b_2, -(a_1 - a_2) r)}. \tag{11}$$

Since both denominator and numerator of the fraction in (11) converge to one as r approaches zero, we have

$$\lim_{r \to 0} P(\tau, r) = A(\tau).$$

(b) We compute the derivative of P with respect to τ:

$$\frac{\partial P}{\partial \tau} = \int_0^r \frac{\partial \pi}{\partial \tau}(\tau, r_1, r - r_1) f(r_1, r) dr_1$$
$$= P(\tau, r) \left[\left(\frac{\dot{A}}{A} - \dot{B}_2 r \right) - (\dot{B}_1 - \dot{B}_2) \frac{\int_0^r r_1 \pi(\tau, r_1, r - r_1) f(r_1, r) dr_1}{\int_0^r \pi(\tau, r_1, r - r_1) f(r_1, r) dr_1} \right], \tag{12}$$

where $\dot{A} = \frac{\partial A}{\partial t}$.
The numerator of the fraction in (12) is positive for all $r > 0$ and can be bounded from above by $r \int_0^r \pi(\tau, r_1, r - r_1) f(r_1, r) dr_1$. Hence the fraction is positive and bounded from above by r, which implies that it converges to zero as $r \to 0$. Since we already know that $P(\tau, r) \to A(\tau)$ for $r \to 0$, we obtain from (12) that

$$\lim_{r \to} \frac{\partial P}{\partial \tau}(\tau, r) = \dot{A}(\tau).$$

(c) In the computation of the derivative $\frac{\partial P}{\partial r}$

$$\frac{\partial P}{\partial r}=\int_0^r \frac{\partial \pi}{\partial r}(\tau,r_1,r-r_1)f(r_1,r)+\pi(\tau,r_1,r-r_1)\frac{\partial f}{\partial r}(r_1,r)dr_1 \tag{13}$$

there are two derivatives which need further computation, $\frac{\partial \pi}{\partial r}$ and $\frac{\partial f}{\partial r}$. Straightforward calculations show that

$$\frac{\partial \pi}{\partial r}(\tau,r_1,r-r_1)=-B_2(\tau)\pi(\tau,r_1,r-r_1) \tag{14}$$

and

$$\begin{aligned}\frac{\partial f}{\partial r}(r_1,r)&=\frac{f_1(r_1)f_2'(r-r_1)}{M(r)}-\frac{f_1(r_1)f_2(r-r_1)}{M^2(r)}M'(r)\\&=f(r_1,r)\left[\frac{f_2'(r-r_1)}{f_2(r-r_1)}-\frac{\int_0^r f_1(s)f_2'(r-s)ds}{\int_0^r f_1(s)f_2(r-s)ds}\right].\end{aligned} \tag{15}$$

Noting that

$$\frac{f_2'(x)}{f_2(x)}=-a_2+(b_2-1)\frac{1}{x},$$

equation (15) can be rewritten as

$$\frac{\partial f}{\partial r}(r_1,r)=f(r_1,r)(b_2-1)\left[\frac{1}{r-r_1}-\frac{\int_0^r \frac{1}{r-s}f_1(s)f_2(r-s)ds}{\int_0^r f_1(s)f_2(r-s)ds}\right]. \tag{16}$$

Substituting (14) and (16) into (13) then yields (after rearrangement)

$$\begin{aligned}\frac{\partial P}{\partial r}=P\Bigg[-B_2+(b_2-1)\Bigg(&\frac{\int_0^r \frac{1}{r-r_1}\pi(\tau,r_1,r-r_1)f(r_1,r)dr_1}{\int_0^r \pi(\tau,r_1,r-r_1)f(r_1,r)dr_1}\\&-\frac{\int_0^r \frac{1}{r-r_1}f_1(r_1)f_2(r-r_1)dr_1}{\int_0^r f_1(r_1)f_2(r-r_1)dr_1}\Bigg)\Bigg].\end{aligned} \tag{17}$$

Let us denote

$$X_1=\frac{\int_0^r \frac{1}{r-r_1}\pi(\tau,r_1,r-r_1)f(r_1,r)dr_1}{\int_0^r \pi(\tau,r_1,r-r_1)f(r_1,r)dr_1},\ X_2=\frac{\int_0^r \frac{1}{r-r_1}f_1(r_1)f_2(r-r_1)dr_1}{\int_0^r f_1(r_1)f_2(r-r_1)dr_1},$$

then,

$$\frac{\partial P}{\partial r}=P(\tau,r)\left[-B_2+(b_2-1)\left(X_1-X_2\right)\right]. \tag{18}$$

The expressions for both X_1 and X_2 can be written in terms of functions ${}_1F_1$:

$$X_1=\frac{1}{r}\frac{b_1+b_2-1}{b_2-1}\frac{{}_1F_1(b_1,b_1+b_2-1,-((B_1-B_2)+(a_1-a_2)r))}{{}_1F_1(b_1,b_1+b_2,-((B_1-B_2)+(a_1-a_2)r))}, \tag{19}$$

and

$$X_2=\frac{1}{r}\frac{b_1+b_2-1}{b_2-1}\frac{{}_1F_1(b_1,b_1+b_2-1,-(a_1-a_2)r)}{{}_1F_1(b_1,b_1+b_2,-(a_1-a_2)r)}. \tag{20}$$

Hence

$$X_1 - X_2 = \frac{\partial P}{\partial r} = \frac{1}{r}\frac{b_1+b_2-1}{b_2-1}\left[\frac{G_1}{G_2} - \frac{G_3}{G_4}\right],$$

where we denoted

$$\begin{aligned} G_1 &= {}_1F_1(b_1, b_1+b_2-1, -((B_1-B_2)+(a_1-a_2))r), \\ G_2 &= {}_1F_1(b_1, b_1+b_2, -((B_1-B_2)+(a_1-a_2))r), \\ G_3 &= {}_1F_1(b_1, b_1+b_2-1, -(a_1-a_2)r), \\ G_4 &= {}_1F_1(b_1, b_1+b_2, -(a_1-a_2)r). \end{aligned} \tag{21}$$

As $G_2G_4 \to 1$ as $r \to 0$, we need to compute $G_1G_4 - G_2G_3$ to be able to compute the limit of (17). Starting from

$$\begin{aligned} G_1 &= 1 - \frac{b_1}{b_1+b_2-1}((B_1-B_2)+(a_1-a_2))r + o(r), \\ G_2 &= 1 - \frac{b_1}{b_1+b_2}((B_1-B_2)+(a_1-a_2))r + o(r), \\ G_3 &= 1 - \frac{b_1}{b_1+b_2-1}(a_1-a_2)r + o(r), \\ G_4 &= 1 - \frac{b_1}{b_1+b_2}(a_1-a_2)r + o(r), \end{aligned} \tag{22}$$

we have

$$G_1G_4 - G_2G_3 \;=\; r\left(-\frac{b_1}{b_1+b_2-1} + \frac{b_1}{b_1+b_2}\right) + o(r), \tag{23}$$

and hence

$$X_1 - X_2 = \frac{b_1+b_2-1}{b_2-1}\frac{1}{G_2G_4}\left[(B_1-B_2)\left(-\frac{b_1}{b_1+b_2-1} + \frac{b_1}{b_1+b_2}\right) + \frac{o(r)}{r}\right]$$

resulting in

$$\lim_{r\to 0} X_1 - X_2 = \frac{b_1+b_2-1}{b_2-1}(B_1-B_2)\left(-\frac{b_1}{b_1+b_2-1} + \frac{b_1}{b_1+b_2}\right).$$

Finally, we can compute the limit of (17) as follows:

$$\begin{aligned} \lim_{r\to 0}\frac{\partial P}{\partial r}(\tau, r) &= \lim_{r\to 0} P(\tau, r)\left[-B_2 + (b_2-1)\left(X_1 - X_2\right)\right] \\ &= A\left[-B_2 + (b_1+b_2-1)(B_1-B_2)\left(-\frac{b_1}{b_1+b_2-1} + \frac{b_1}{b_1+b_2}\right)\right] \\ &= -A\left[\frac{b_1}{b_1+b_2}B_1 + \frac{b_2}{b_1+b_2}B_2\right]. \end{aligned}$$

(d) We show that there is a finite limit of $\frac{\partial^2 P}{\partial r^2}(\tau, r)$ as $r \to 0$, from which the boundedness follows.
From (17) we know that

$$\frac{\partial^2 P}{\partial r^2} = \frac{\partial P}{\partial r}\left[-B_2 + (b_2 - 1)(X_1 - X_2)\right] + P\frac{\partial\left[-B_2 + (b_2 - 1)(X_1 - X_2)\right]}{\partial r}.$$

From the definition of X_1 and X_2 and from their limits, it follows that it is sufficient to show the existence of the finite limit of $\frac{\partial}{\partial r}\left(\frac{1}{r}F(r)\right)$ for $r \to 0+$, where

$$F(r) = \frac{G_1(r)}{G_2(r)} - \frac{G_3(r)}{G_4(r)}. \qquad (24)$$

With $F(r)$ written by means of a series expansion $F(r) = \sum_{k=0}^{\infty} a_k r^k$, the condition $a_0 = 0$ is sufficient for boundedness of the term $\frac{\partial}{\partial r}\left(\frac{1}{r}F(r)\right)$ in the neighbourhood of $r = 0$, which holds for (24).

■

This leads us to the main result of this paper.

Theorem 3.3 *Under the hypotheses of theorem 3.1, there is no one-factor interest rate model for which the averaged bond prices satisfy the PDE up to the boundary* $r = 0$.

Proof. By taking the limit $r \to 0$ in the PDE (2), we know from the previous theorem that for all $\tau > 0$

$$-\dot{A}(\tau) + \alpha(r=0)(-A(\tau))\left(\frac{b_1}{b_1 + b_2}B_1(\tau) + \frac{b_2}{b_1 + b_2}B_2(\tau)\right) = 0.$$

From this we calculate the value of the function α for $r = 0$:

$$\alpha(r=0) = -\frac{\dot{A}(\tau)}{A(\tau)}\frac{1}{\frac{b_1 B_1(\tau)}{b_1+b_2} + \frac{b_2 B_2(\tau)}{b_1+b_2}} = -\frac{\dot{A}(\tau)}{A(\tau)}\frac{b_1 + b_2}{b_1 B_1(\tau) + b_2 B_2(\tau)}.$$

This means that

$$-\frac{\dot{A}(\tau)}{A(\tau)}\frac{b_1 + b_2}{b_1 B_1(\tau) + b_2 B_2(\tau)} = K_1 \qquad (25)$$

is a constant, independent of τ.

Now we use the fact that the function $A(\tau)$ in the two-factor CIR model can be written as $A(\tau) = A_1(\tau)A_2(\tau)$, where $A_1(\tau)$ and $A_2(\tau)$ are functions of the original CIR models, corresponding to each of the equations for r_1 and r_2. Hence they satisfy

$$\dot{A}_i(\tau) = \kappa_i \theta_i A_i(\tau) B_i(\tau) \quad (i = 1, 2)$$

and so we get

$$\frac{\dot{A}(\tau)}{A(\tau)} = \frac{\dot{A_1}(\tau)A_2(\tau) + A_1(\tau)\dot{A_2}(\tau)}{A_1(\tau)A_2(\tau)} = \frac{\dot{A}_1(\tau)}{A_1(\tau)} + \frac{\dot{A_2}(\tau)}{A_2(\tau)} = \kappa_1\theta_1 B_1(\tau) + \kappa_2\theta_2 B_2(\tau).$$

The expression in (25) can be rewritten as

$$K_1 = -\frac{\dot{A}(\tau)}{A(\tau)}\frac{b_1+b_2}{b_1B_1(\tau)+b_2B_2(\tau)} = -(\kappa_1\theta_1B_1(\tau)+\kappa_2\theta_2B_2(\tau))\frac{b_1+b_2}{b_1B_1(\tau)+b_2B_2(\tau)}.$$

Since b_1+b_2 is constant, the important part is the following fraction, which has to be equal to some constant K:

$$\frac{\kappa_1\theta_1B_1(\tau)+\kappa_2\theta_2B_2(\tau)}{b_1B_1(\tau)+b_2B_2(\tau)} = K.$$

It implies that

$$\kappa_1\theta_1B_1(\tau)+\kappa_2\theta_2B_2(\tau) = K(b_1B_1(\tau)+b_2B_2(\tau))$$

and so

$$(\kappa_1\theta_1 - Kb_1)B_1(\tau) = (Kb_2 - \kappa_2\theta_2)B_2(\tau)$$

for each $\tau > 0$. This is only possible in two cases:

1. $\kappa_1\theta_1 - Kb_1 = 0$ and $Kb_2 - \kappa_2\theta_2 = 0$,
2. $B_1(\tau) = cB_2(\tau)$ for some constant c.

Let us look at each of these possibilities.

1. The same constant K appears in both conditions. From the first one we get $K = \frac{\kappa_1\theta_1}{b_1}$ and by substituting the value of $b_1 = \frac{2\kappa_1\theta_1}{\sigma_1^2}$, we obtain $K = \frac{\sigma_1^2}{2}$. In the same way, from the second equality we obtain $K = \frac{\sigma_2^2}{2}$. However, we started from the hypothesis that $\sigma_1^2 \neq \sigma_2^2$, and thus we find a contradiction.

2. We recall the equations for B_1 and B_2 from the CIR model:

$$-\dot{B}_i(\tau) = (\kappa_i + \lambda_i\sigma_i)B_i(\tau) + \frac{1}{2}\sigma_i^2B_i(\tau)^2 - 1. \tag{26}$$

If $B_1(\tau) = cB_2(\tau)$, it follows that

$$c\left[(\kappa_2+\lambda_2\sigma_2)B_2(\tau) + \frac{1}{2}\sigma^2B_2(\tau)^2 - 1\right] = (\kappa_1+\lambda_1\sigma_1)B_1(\tau) + \frac{1}{2}\sigma_1^2B_1(\tau)^2 - 1$$

for all $\tau > 0$. By continuity, the equality also holds for the limit $\tau = 0+$. Taking this limit, we get $c = 1$, and hence the functions $B_1(\tau)$ and $B_2(\tau)$ coincide. We denote this function by $B(\tau)$. Subtracting the two equations in (26), we obtain:

$$\left[-(\kappa_1+\lambda_1\sigma_1)+(\kappa_2+\lambda_1\sigma_1)\right]B(\tau) + \left[-\frac{1}{2}\sigma_1^2+\frac{1}{2}\sigma_2^2\right]B^2(\tau) = 0$$

and dividing by $B(\tau)$ (which is nonzero)

$$\left[-(\kappa_1+\lambda_1\sigma_1)+(\kappa_2+\lambda_1\sigma_1)\right] - \frac{1}{2}\left[\sigma_2^2-\sigma_1^2\right]B(\tau) = 0.$$

Since $\sigma_1 \neq \sigma_2$, this implies that $B(\tau)$ is a constant function, which is a contradiction.

Since both possibilities lead to a contradiction, the theorem is proved. ∎

4. CONCLUSION

In this paper, we considered two-factor Cox-Ingersoll-Ross models for interest rates and averaged bond prices with respect to the asymptotic distribution of the short rate processes, conditioned on the observable short rate level. Such averaged values are functions of the maturity and of the short rate, just as the solutions of one-factor models. Hence we studied the question, whether there would be a one-factor model yielding the same bond prices as those obtained by averaging in the two-factor Cox-Ingersoll-Ross model. We proved that the answer is negative. In the future, we plan to study this question also for other two-factor models.

References

M. Abramovitz and I. Stegun. *Handbook on Mathematical functions*. National Bureau of Standards, 1972.

T.G. Anderson and J. Lund. The short rate diffusion revisited: A tour guided by the efficient method of moments. Working paper, 1996.

M.J. Brennan and E.S. Schwartz. An equilibrium model of bond pricing and a test of market efficiency. *The Journal of Financial and Quantitative Analysis*, 17(3):301–329, 1982.

K.C. Chan, G.A. Karolyi, F.A. Longstaff, and A.B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*, 47(3):1209–1227, 1992.

C. Christiansen. Multivariate term structure models with levels and heteroskedasticity effects. Centre for Analytical Finance, University of Aarhus Working paper series, No. 129, 2002.

T. Corzo Santamaria and J.G. Biscarri. Nonparametric estimation of convergence of interest rates: Effects on bond pricing. *Spanish Economic Review*, 7(3):167–190, 2005.

T. Corzo Santamaria and E.S. Schwartz. Convergence within the European Union: Evidence from interest rates. *Economic Notes*, 29:243–266, 2000.

P. Cotton, J.-P. Fouque, G. Papanicolaou, and R. Sircar. Stochastic volatility corrections for interest rate derivatives. *Mathematical Finance*, 14(2):173–200, 2004.

J.C. Cox, J.E. Ingersoll, and S.A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

H.G. Fong and O.A. Vasicek. Fixed-income volatility management. *Journal of Portfolio Management*, 17(4):41–46, 1991.

J.-P. Fouque, G. Papanicolaou, K.R. Sircar, and K. Solna. Singular perturbations in option pricing. *SIAM Journal on Applied Mathematics*, 63(5):1648–1665, 2003.

Y.K. Kwok. *Mathematical models of financial derivatives*. Springer - Verlag, 1998.

S.M. Schaefer and E.S. Schwartz. A two-factor model of the term structure: An approximate analytical solution. *The Journal of Financial and Quantitative Analysis*, 19(4):413–424, 1984.

O.A. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

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 "5th Actuarial and Financial Mathematics Day" (9 februari 2007, Prof. M. Vanmaele)

De "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 niet alleen in België maar ook daarbuiten. Naast twee gastsprekers kwamen doctoraatsstudenten, postdocs en mensen uit de bedrijfswereld aan bod. Voor deze editie werd geopteerd voor wat minder sprekers om hen langer aan het woord te laten en om tijd te laten voor discussie. We danken hierbij in het bijzonder de discussanten als gangmakers van deze discussies.
In deze publicatie vindt u een neerslag van een aantal voorgestelde onderwerpen. De 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, "variance reduction" methoden, hedging en equilibrium vraagstukken, exotische opties en rentemodellen.