Accepted Manuscript

Insurance valuation: A computable multi-period cost-of-capital approach

Hampus Engsner, Mathias Lindholm, Filip Lindskog

Please cite this article as: Engsner, H., Lindholm, M., Lindskog, F., Insurance valuation: A computable multi-period cost-of-capital approach. *Insurance: Mathematics and Economics* (2016), http://dx.doi.org/10.1016/j.insmatheco.2016.12.002

INSURANCE VALUATION: A COMPUTABLE MULTI-PERIOD COST-OF-CAPITAL APPROACH

HAMPUS ENGSNER, MATHIAS LINDHOLM, AND FILIP LINDSKOG

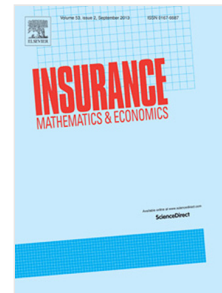ABSTRACT. We present an approach to market-consistent multi-period valuation of insurance liability cash flows based on a two-stage valuation procedure. First, a portfolio of traded financial instrument aimed at replicating the liability cash flow is fixed. Then the residual cash flow is managed by repeated one-period replication using only cash funds. The latter part takes capital requirements and costs into account, as well as limited liability and risk averseness of capital providers. The cost-of-capital margin is the value of the residual cash flow. We set up a general framework for the cost-of-capital margin and relate it to dynamic risk measurement. Moreover, we present explicit formulas and properties of the cost-of-capital margin under further assumptions on the model for the liability cash flow and on the conditional risk measures and utility functions. Finally, we highlight computational aspects of the cost-of-capital margin, and related quantities, in terms of an example from life insurance.

Keywords: valuation of insurance liabilities, multi-period valuation, market-consistent valuation, cost of capital, risk margin, dynamic risk measurement

JEL Classification: G22, G11, G28

1. INTRODUCTION

The current solvency regulatory framework Solvency II emphasizes market-consistent valuation of liabilities; it is explicitly stated that liabilities should be "valued at the amount for which they could be ... transferred or settled ... between knowledgeable and willing parties in an arm's length transaction". Solvency assessment of an insurance company is based on future net values of assets and liabilities, and market-consistent valuation enables solvency assessments that take dependence between future values of assets and liabilities into account. Moreover, current regulatory frameworks emphasize risk measurement over a one-year period. In particular, at any given time, the whole liability cash flow is taken into account in terms of the cash flow during the next one-year period and the market-consistent value at the end of the one-year period of the remaining liability cash flow. However, liability cash flows are typically not replicable by financial instruments. Therefore, the contribution to the liability value from the residual cash flow resulting from imperfect replication must be determined.

Given an aggregate liability cash flow of an insurance company, portfolios may be formed that generate cash flows with expected values matching that

of the liability cash flow. The traditional actuarial practice of reserving provides an example of such a portfolio consisting of default-free bonds. In case of dependence between the liability cash flow and market values of financial instruments, more sophisticated replicating portfolios may be more suitable. However, the mismatch between the cash flow of such a portfolio and that of the original liability cash flow is typically substantial. The residual liability cash flow must be handled throughout the life of the liability cash flow by making sure that sufficient additional capital is available at all times. Capital providers, such as share holders, require compensation for providing buffer capital, which should be taken into account in the liability valuation. In particular, capital costs should be accounted for.

In Solvency II, the so-called technical provisions correspond to the aggregate liability value and is defined as the sum of a best estimate, corresponding to a discounted actuarial fair value, and a so-called risk margin aimed at capturing capital costs. Unfortunately, the risk margin in the current regulatory framework lacks a proper definition and theoretical foundation, and different approximation formulas for this ill-defined object have been suggested. Criticism of the risk margin and suggestions for better notions of cost-of-capital margins or market-value margins are found in e.g. [11], [13], [17], [20] and [24].

This paper addresses valuation of an aggregate liability cash flow of an insurance company, although the problem and our suggested solutions apply to liability valuation in other contexts as well. We present an approach that, in many aspects, is similar to current practice and has wide-ranging applicability. The approach we propose is inspired by [11] and [17], where the cost-of-capital margin for valuing aggregate liability cash flows is analyzed. The framework for liability cash flow valuation presented in [11] combines financial replication arguments with cost-of-capital considerations. Our proposed framework is on the one hand a generalization of that framework and with more attention paid to mathematical details. On the other hand, for the repeated one-period replication of the residual cash flow, we severely restrict the allowed replication instruments compared to [11]. Such a simplification of the problem allows us to derive much stronger results which in turn yields a framework that can be easily adopted, and it avoids many of the computational difficulties highlighted in [17] without sacrificing conceptual consistency. From a practical perspective, it leads to an approach to valuation of liability cash flows that does not rely heavily on subjective choices of joint dynamics for market prices of possible replication instruments. The use of financial valuation principles in insurance is inevitable given the principle of market-consistent valuation of liabilities, in particular for liability cash flows with long durations and products with guarantees. For more on financial and actuarial valuation of insurance liabilities aimed at solvency assessment, see [25].

Conditional monetary risk measures and utility functions are important basic building blocks in the approach to liability cash flow valuation considered here. Extensions from static, or one-period, risk measurement to dynamic risk measurement has been studied extensively for more than a decade following the seminal paper [2] on one-period risk measurement, see

e.g. [3], [4], [6], [16], [19] and [23]. Much of the analysis of dynamic risk measurement has focused on dynamic measurement of risk corresponding to a single cash flow, such as the cash flow of a derivative payoff, at a fixed future time. In [20] a market-value margin for valuation of insurance liabilities is presented based on multi-period mean-variance hedging. However, this valuation framework is not directly applicable to the problem we consider since the liability cash flow considered in [20] occur at a terminal time whereas we consider cash flows at all times up to a terminal time. For the problem we consider, there is no natural way to roll cash flows forward and thereby reducing the dynamic risk measurement problem to a considerably simpler problem. The frameworks for dynamic risk measurement developed in [3] and [4] are well-suited to handle liability valuation problems of the type we consider. However, we do not want to restrict the liability cash flows to bounded stochastic processes. Moreover, the liability valuation problem we consider corresponds to repeated one-period replication rather than truly multi-period replication. Another important aspect is that we assume that the capital provider has limited liability and that causes the cost-of-capital margin to lack the convexity properties that are essential for the so-called risk-adjusted values analyzed in [3] and similarly for the dynamic risk measures analyzed in [4]. Risk measurement for multi-period income streams are studied in [14] and [15], where the dynamic risk measurement problem is formulated as a stochastic optimization problem. There, computational aspects of multi-period risk measurement are clarified and illustrated. Although our approach to liability valuation differ substantially from that in [14] and [15], computability is an essential feature.

This paper is organized as follows. Section 2 gives a nontechnical derivation of the cost-of-capital margin, as we believe that it should be defined, by economic arguments.

Section 3 presents a mathematical framework that allows the cost-of-capital margin to be defined rigorously, and establishes its fundamental properties. We also describe how the cost-of-capital margin is related to conditional monetary utility functions in the sense of [4], showing that the cost-of-capital margin is conceptually consistent with risk-adjusted values and dynamic monetary utility functions as defined in [3] and [4]. There is however a major difference. The limited liability property of capital providers is an essential ingredient in our definition of the cost-of-capital margin and may cause violation of concavity/convexity properties. We define the cost-of-capital margin in terms of repeated one-period replication similar to [11] and allow capital requirements to be given in terms of conditional versions of nonconvex risk measures such as Value-at-Risk which is the current industry practice for insurance markets subject to the Solvency II regulation. Therefore, convexity properties are not assumed and not essential to us. Time-consistency is however an essential property. For the cost-of-capital margin, this property essentially follows immediately from the definition. Towards the end of Section 3 we consider a family of conditional risk measures that include commonly used risk measures such as Value-at-Risk and spectral risk measures, and we show that this family of risk measures are particularly useful for ensuring stronger properties and

explicit formulas for the cost-of-capital margin when the liability cash flow is restricted to certain families of stochastic processes. It is well known, see e.g. [5] and [19], that conditional or dynamic versions of Value-at-Risk and spectral risk measures are not time-consistent when applied to time periods of varying lengths. In our setting, only repeated conditional single-period risk measurement appears. Therefore, time-inconsistency of risk measurement over time periods of varying lengths does not cause problems for the time-consistency of the cost-of-capital margin.

Section 4 considers specific models for the liability cash flow and the filtration representing the flow of information over time about the remaining cash flow until complete runoff of the liability. Specifically, we consider models of autoregressive type and Gaussian models. We show that when combined with the general framework presented in Section 3, these models allow for explicit formulas and stronger results concerning the effects of properties of the chosen filtration. We believe that the explicit expressions presented here constitute candidates for standard formulas for cost-of-capital margin computation that may be adopted in improved future solvency regulation.

Finally, Section 5 presents a life-insurance example that illustrates features of the cost-of-capital margin and clarifies computational aspects.

2. The cost-of-capital margin

In this section we derive the cost-of-capital margin without mathematical details, they are found in Section 3.

We consider time periods (years) $1, \ldots, T$, corresponding time points $0, 1, \ldots, T$, and a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = (\mathcal{F}_t)_{t=0}^{T}$ with $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \cdots \subseteq \mathcal{F}_T = \mathcal{F}$.

A liability cash flow corresponds to an $\mathbb{F}$-adapted stochastic process $X^o = (X_t^o)_{t=1}^{T}$ interpreted as a cash flow from an aggregate insurance liability in runoff. Our aim is to give a precise meaning to the market-consistent value of the liability by taking capital costs into account, and provide results that allow this value to be computed.

When the value of an insurance liability cash flow includes capital costs from capital requirements based on future values of both assets and liabilities, the liability value depends on the future values of all assets, including assets held for investment purpose only. In particular, two companies with identical liability cash flows would assign different market-consistent values to the two identical cash flows. This has undesired implications. Instead, as is done in e.g. [11] and prescribed by EIOPA, see [9, Article 38], we take the point of view that an aggregate liability cash flow should be valued by considering a hypothetical transfer of the liability to a separate entity, a so-called reference undertaking, whose assets have the sole purpose of matching the value of the liability as well as possible.

We will give a meaning to the liability value by a particular two-stage valuation procedure: the first stage corresponds to choosing a replicating portfolio of traded financial instruments, the second stage corresponds to managing the residual cash flow from imperfect replication in the first stage.

At time 0, a portfolio is purchased with the aim of generating a cash flow replicating the cash flow X^o. This static replicating portfolio has a market

price π and generates the cash flow $X^s = (X_t^s)_{t=1}^T$. We use the wording "static" in order to emphasize that, for the purpose of valuation, it is a portfolio strategy that is fixed throughout the life of the liability cash flow. However, the replicating portfolio may be "dynamic" in the sense that its cash flow depends on events not known at time 0. This is completely in line with pricing a financial derivative in terms of the initial market price of a self-financing hedging strategy. If X^o is independent of financial asset prices, then the canonical example is a portfolio of zero-coupon bonds generating the cash flow $X^s = \mathbb{E}[X^o] = (\mathbb{E}[X_t^o])_{t=1}^T$.

The value of the original liability is defined as the sum of the market price π of the replicating portfolio and the value $V_0(X)$ of the residual cash flow $X := X^o - X^s$ from repeated one-period replication using only cash funds provided by a capital provider with limited liability requiring compensation for capital costs. That is, the cash flow X will be re-valued at all times $1, \dots, T$. We call the value $V_0(X)$ of the residual cash flow the cost-of-capital margin. Note that we below refer also to $V_t(X)$ as the cost-of-capital margin for all $t = 0, \dots, T$. We emphasize that repeated one-period replication is done with cash only. Allowing for repeated one-period replication using a mix of assets such as bonds with short time to maturity inevitably makes the value of the liability cash flow depend on subjective views on the development over time of spot rates over different time horizons.

Next we will present economic arguments that lead to a recursion defining $V_t(X)$ in terms of X_{t+1} and $V_{t+1}(X)$, capital requirements and the acceptability condition of the capital provider. At time t, the insurance company is required to hold the capital $R_t(-X_{t+1} - V_{t+1}(X))$, where R_t is a conditional monetary risk measure, see Definition 1, quantifying the risk from liability cash flow during year $t+1$ and the value at time $t+1$ of the remaining residual cash flow $(X_{t+2}, \dots, X_T)$. Also at time t, the capital $V_t(X)$ matching precisely the value of the residual cash flow at that time is available. The capital provider is asked to provide the missing required capital

$$(1) \qquad C_t := R_t(-X_{t+1} - V_{t+1}(X)) - V_t(X).$$

If the capital provider accepts providing C_t at time t, then at time $t+1$ the amount X_{t+1} is paid and the remaining capital is $R_t(-X_{t+1} - V_{t+1}(X)) - X_{t+1}$. If this amount exceeds the value $V_{t+1}(X)$ of the liability at time $t+1$, then the capital provider collects the excess capital as a compensation for providing the buffer capital C_t at time t. Moreover, the capital provider has limited liability: if

$$R_t(-X_{t+1} - V_{t+1}(X)) - X_{t+1} - V_{t+1}(X) < 0,$$

then the capital provider has no obligation to provide further capital to offset the deficit.

The capital provider's acceptability condition at time t is expressed in terms of a conditional utility function U_t, see Definition 2, and a value $\eta_t \geq 0$ quantifying the size of the intended compensation to the capital provider for making capital available:

$$(2) \qquad U_t\big(\big(R_t(-X_{t+1} - V_{t+1}(X)) - X_{t+1} - V_{t+1}(X)\big)_+\big) \geq (1 + \eta_t)C_t,$$

where $x_+ = \max(x, 0)$. U_t may be chosen as the conditional expectation although alternatives that take the risk aversion of the capital provider into account may be more appropriate.

Combining (1) and (2) now gives, with $Y_{t+1} := X_{t+1} + V_{t+1}(X)$,

$$V_t(X) \geq R_t(-Y_{t+1}) - \frac{1}{1 + \eta_t} U_t\big((R_t(-Y_{t+1}) - Y_{t+1})_+\big).$$

If the inequality above is strict, then the capital provider obtains a better-than-required investment opportunity at the expense of policy holders who are obliged to pay higher-than-needed premiums: the capital $V_t(X)$ belongs to the policy holders of contracts that at that time may still cause claims, not to the capital provider. Therefore, we define the value of the cash flow as the smallest value for which the capital provider finds the investment opportunity acceptable. That is, we replace the inequality above by an equality:

$$(3) \qquad V_t(X) := R_t(-Y_{t+1}) - \frac{1}{1 + \eta_t} U_t\big((R_t(-Y_{t+1}) - Y_{t+1})_+\big).$$

Recall that the market-consistent value we assign to the original cash flow is the sum of the market price π of a replicating portfolio set up at time 0 and the cost-of-capital margin $V_0(X)$. Notice that if perfect initial replication is possible, then $X = 0$ as there is no residual cash flow, and consequently $V_0(X) = 0$ since no capital funds for repeated one-period replication are needed. In this case the market-consistent value is simply the market price of the perfect hedge of the liability cash flow. However, when perfect replication is impossible, a poor choice of replicating portfolio may lead to a residual cash flow that requires much capital from the capital provider, causing the market-consistent value to become unreasonably high. In particular, for this approach to market-consistent valuation of insurance liabilities to be sound, in the sense that policy holders are not over-charged due to poor asset-liability management, the insurance regulator must specify what approaches to imperfect replication of liability cash flows are acceptable.

3. The valuation framework

We consider time periods $1, \ldots, T$, corresponding time points $0, 1, \ldots, T$, and a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = (\mathcal{F}_t)_{t=0}^T$ with $\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \cdots \subseteq \mathcal{F}_T = \mathcal{F}$. Let $L^0(\mathcal{F}_t) := L^0(\Omega, \mathcal{F}_t, \mathbb{P})$ denote the vector space of all real-valued $\mathcal{F}_t$-measurable random variables, and let $L^0_+(\mathcal{F}_t)$ be the subset of $L^0(\mathcal{F}_t)$ of random variables taking values in $[0, \infty)$. $L^0(\mathcal{F}_0)$ is simply the set of constants, i.e. $\mathbb{R}$. For $p \in (0, \infty)$, let $L^p(\mathcal{F}_t) := \{Y \in L^0(\mathcal{F}_t) : \mathbb{E}[|Y|^p] < \infty\}$ and let $L^p_+(\mathcal{F}_t)$ denotes the subset of $L^p(\mathcal{F}_t)$ of random variables taking values in $[0, \infty)$. Finally, $L^\infty(\mathcal{F}_t)$ consists of the essentially bounded $\mathcal{F}_t$-measurable random variables: $Y \in L^0(\mathcal{F}_t)$ such that

$$\inf\{r > 0 : \mathbb{P}(\omega \in \Omega : |Y(\omega)| > r) = 0\} < \infty.$$

We say that two random variables are equal if they coincide $\mathbb{P}$-almost surely (a.s.). All equalities and inequalities between random variables are interpreted in the $\mathbb{P}$-a.s. sense.

In order to determine the cost-of-capital margin we consider conditional monetary risk measures R_t and conditional monetary utility functions U_t. We express values and cash flows via a numéraire which we take to be a money market account that pays no interest, where money can be deposited and later withdrawn. We have no need for and do not assume risk-free borrowing. Choosing the numéraire to be a money market account paying stochastic interest rates would not pose mathematical difficulties, but would force us to pay more attention to the interpretation of the cash flows in terms of this numéraire.

By a dynamic monetary risk measure $R := (R_t)_{t=0}^{T-1}$ quantifying one-period capital requirements we mean the following:

Definition 1. *For $p \in [0, \infty]$, a dynamic monetary risk measure $(R_t)_{t=0}^{T-1}$ is a sequence of mappings $R_t : L^p(\mathcal{F}_{t+1}) \to L^p(\mathcal{F}_t)$ satisfying*

(4)   *if $\lambda \in L^p(\mathcal{F}_t)$ and $Y \in L^p(\mathcal{F}_{t+1})$, then $R_t(Y + \lambda) = R_t(Y) - \lambda$,*

(5)   *if $Y, \widetilde{Y} \in L^p(\mathcal{F}_{t+1})$ and $Y \le \widetilde{Y}$, then $R_t(Y) \ge R_t(\widetilde{Y})$,*

(6) $R_t(0) = 0.$

By a dynamic monetary utility function $U := (U_t)_{t=0}^{T-1}$ quantifying one-period acceptability for capital providers we mean the following:

Definition 2. *For $p \in [0, \infty]$, a dynamic monetary utility function $(U_t)_{t=0}^{T-1}$ is a sequence of mappings $U_t : L_+^p(\mathcal{F}_{t+1}) \to L_+^p(\mathcal{F}_t)$ satisfying*

(7)   *if $\lambda \in L_+^p(\mathcal{F}_t)$ and $Y \in L_+^p(\mathcal{F}_{t+1})$, then $U_t(Y + \lambda) = U_t(Y) + \lambda$,*

(8)   *if $Y, \widetilde{Y} \in L_+^p(\mathcal{F}_{t+1})$ and $Y \le \widetilde{Y}$, then $U_t(Y) \le U_t(\widetilde{Y})$,*

(9) $U_t(0) = 0.$

Notice from (3) that, for the purpose of defining the cost-of-capital margin, it is sufficient to only consider mappings U_t whose argument is nonnegative.

Remark 1. *If $(R_t)_{t=0}^{T-1}$ is a dynamic monetary risk measure in the sense of Definition 1, then $(\widetilde{U}_t)_{t=0}^{T-1} := (-R_t)_{t=0}^{T-1}$ has the properties*

(10)   *if $\lambda \in L^p(\mathcal{F}_t)$ and $Y \in L^p(\mathcal{F}_{t+1})$, then $\widetilde{U}_t(Y + \lambda) = \widetilde{U}_t(Y) + \lambda$,*

(11)   *if $Y, \widetilde{Y} \in L^p(\mathcal{F}_{t+1})$ and $Y \le \widetilde{Y}$, then $\widetilde{U}_t(Y) \le \widetilde{U}_t(\widetilde{Y})$,*

(12) $\widetilde{U}_t(0) = 0.$

Properties (11) and (12) together imply $\widetilde{U}_t(Y) \in L_+^p(\mathcal{F}_t)$ if $Y \in L_+^p(\mathcal{F}_{t+1})$. In particular, if, for all t, $\widetilde{U}_t^+$ denotes the restriction of $\widetilde{U}_t$ to $L_+^p(\mathcal{F}_{t+1})$, then $(\widetilde{U}_t^+)_{t=0}^{T-1}$ is a dynamic monetary utility function in the sense of Definition 2.

Remark 2. *Classic expected utility theory gives an acceptability condition of the form $\mathbb{E}[u_t(Y) \mid \mathcal{F}_t] \ge u_t(C_t)$, where u_t is an $\mathcal{F}_t$-measurable, increasing and concave function, and, in our case $Y := (R_t(-Y_{t+1}) - Y_{t+1})_+$. Defining $\widetilde{U}_t(Y) := u_t^{-1}(\mathbb{E}[u_t(Y) \mid \mathcal{F}_t])$ (the certainty equivalent) and setting $\eta_t := 0$ shows that the condition $\mathbb{E}[u_t(Y) \mid \mathcal{F}_t] \ge u_t(C_t)$ is a special case of (2).*

Notice that $\widetilde{U}_t$ satisfies (11) *and* (12) *by construction. In general, $\widetilde{U}_t$ does not satisfy* (10). *However, if u_t is the exponential utility function $u_t(x) = -\tau_t \exp\{-x/\tau_t\}$, $0 < \tau_t \in L^1_+(\mathcal{F}_t)$, then $\widetilde{U}_t(Y) = -\tau_t \log \mathbb{E}\left[\exp\{-Y/\tau_t\} \mid \mathcal{F}_t\right]$ which satisfies* (10). *Hence, for $p = 0$, $(\widetilde{U}_t)_{t=0}^{T-1}$ is a dynamic monetary utility function in the sense of Definition 2.*

In [12], *properties of the functional $Y \mapsto u^{-1}(\mathbb{E}\left[u(Y)\right])$ are determined and related to properties of the increasing and concave function u. In particular, from* [12] *it is clear that the property* (10) *is an exception.*

There is a vast literature on expected utility theory and how it appears from axioms of rational decision making, see e.g. [18]. *In* [26], *the consequences of replacing the so-called Independence Axiom by an axiom termed Dual Independence was presented, leading to the so-called Dual Theory of Choice Under Risk. Whereas the traditional independence axiom leads to preference relations expressed in terms of expected utilities, under the Dual Theory preference relations can be phrased in terms of utility functionals having (when modified to a conditional setting such as the one considered here) the property* (10) *(in* [26], *the utility functionals are defined on a space of bounded random variables). For further details we refer to Section 3.3 and in particular Remark 4.*

The following proposition provides the basis for defining the cost-of-capital margin in (3) rigorously.

Proposition 1. *Fix $p \in [0, \infty]$. Let $(R_t)_{t=0}^{T-1}$ be given by Definition 1, let $(U_t)_{t=0}^{T-1}$ be given by Definition 2, and let $\eta_t \in L^0_+(\mathcal{F}_t)$.*
(i)

$$(13) \qquad W_t(Y) := R_t(-Y) - \frac{1}{1+\eta_t} U_t\big((R_t(-Y) - Y)_+\big)$$

is a mapping from $L^p(\mathcal{F}_{t+1})$ to $L^p(\mathcal{F}_t)$ having the properties

$$(14) \qquad \text{if } \lambda \in L^p(\mathcal{F}_t) \text{ and } Y \in L^p(\mathcal{F}_{t+1}), \text{ then } W_t(Y + \lambda) = W_t(Y) + \lambda,$$

$$(15) \qquad \text{if } Y, \widetilde{Y} \in L^p(\mathcal{F}_{t+1}) \text{ and } Y \leq \widetilde{Y}, \text{ then } W_t(Y) \leq W_t(\widetilde{Y}),$$

$$(16) \qquad W_t(0) = 0.$$

(ii) Let $(X_t)_{t=1}^{T}$ be an $\mathbb{F}$-adapted cash flow with $X_t \in L^p(\mathcal{F}_t)$ for every t. The cost-of-capital margin $V_t(X)$ in (3) *satisfies*

$$(17) \qquad V_t(X) = W_t \circ \cdots \circ W_{T-1}(X_{t+1} + \cdots + X_T),$$

where $W_t \circ \cdots \circ W_{T-1}$ denotes the composition of mappings $W_t, \ldots, W_{T-1}$.

Proof of Proposition 1. Since $(1 + \eta_t)^{-1}$ is $\mathcal{F}_t$-measurable and takes values in $(0, 1]$, it follows directly from the definitions of R_t and U_t that W_t is a mapping from $L^p(\mathcal{F}_{t+1})$ to $L^p(\mathcal{F}_t)$. The properties (14) and (16) for W_t follow immediately from the corresponding properties of R_t and U_t. It remains to verify property (15) for W_t. Take $Y_{t+1} \leq \widetilde{Y}_{t+1}$ in $L^p(\mathcal{F}_{t+1})$. Then

$$(R_t(-\widetilde{Y}_{t+1}) - \widetilde{Y}_{t+1})_+ \leq (R_t(-Y_{t+1}) - \widetilde{Y}_{t+1})_+ + R_t(-\widetilde{Y}_{t+1}) - R_t(-Y_{t+1})$$
$$\leq (R_t(-Y_{t+1}) - Y_{t+1})_+ + R_t(-\widetilde{Y}_{t+1}) - R_t(-Y_{t+1}).$$

Since $R_t(-\widetilde{Y}_{t+1}) - R_t(-Y_{t+1}) \in L_+^p(\mathcal{F}_t)$, (7)-(9) together imply that

$$U_t((R_t(-\widetilde{Y}_{t+1}) - \widetilde{Y}_{t+1})_+) - U_t((R_t(-Y_{t+1}) - Y_{t+1})_+)$$
$$\leq R_t(-\widetilde{Y}_{t+1}) - R_t(-Y_{t+1})$$

which further implies

$$W_t(\widetilde{Y}_{t+1}) - W_t(Y_{t+1}) \geq (R_t(-\widetilde{Y}_{t+1}) - R_t(-Y_{t+1}))\frac{\eta_t}{1+\eta_t} \geq 0$$

and verifies property (15). Finally, we verify the representation (17) of $V_t(X)$ in terms of $W_t, \ldots, W_{T-1}$ and $X_{t+1} + \cdots + X_T$. If $X_s \in L^p(\mathcal{F}_s)$ for $s = t+1, \ldots, T$, then $X_{t+1} + \cdots + X_T \in L^p(\mathcal{F}_T)$ and the right-hand side in (17) is well-defined. Repeated application of (14) now verifies the representation (17). $\qquad\square$

Definition 3. *Fix $p \in [0, \infty]$. Let $(R_t)_{t=0}^{T-1}$ by given by Definition 1, let $(U_t)_{t=0}^{T-1}$ be given by Definition 2, and let $(W_t)_{t=0}^{T-1}$ be given by (13). Let $(X_t)_{t=1}^T$ be an $\mathbb{F}$-adapted cash flow with $X_t \in L^p(\mathcal{F}_t)$ for every t. We define the cost-of-capital margins $V_t(X)$ as*

$$(18) \qquad V_t(X) := W_t \circ \cdots \circ W_{T-1}(X_{t+1} + \cdots + X_T), \quad t = 0, \ldots, T-1,$$

and $V_T(X) := 0$.

Notice that we may express (18) as $V_t(X) := W_t(X_{t+1} + V_{t+1}(X)) = W_t(Y_{t+1})$.

Remark 3. *Property (14), which is essential for defining the cost-of-capital margin in terms of compositions of W_ts in (18), is inherited from the corresponding property (4) of the R_ts. The corresponding property (7) of the U_ts is used only to show the monotonicity property (15). In fact, a sufficient requirement would be a milder version of (7) with the equality replaced by the inequality $U_t(Y + \lambda) \leq U_t(Y) + \lambda$.*

Proposition 2. *Let $X, \widetilde{X}$ be $\mathbb{F}$-adapted cash flows with $X_t, \widetilde{X}_t \in L^p(\mathcal{F}_t)$ for every t.*
(i) Let b be a T-dimensional vector with components in $L^p(\mathcal{F}_t)$, and let $X_u \leq \widetilde{X}_u$ for each u. Then, for every $t < T$,

$$V_t(0) = 0, \quad V_t(X + b) = V_t(X) + \sum_{u=t+1}^T b_t, \quad V_t(X) \leq V_t(\widetilde{X}).$$

(ii) The cost-of-capital margins are time consistent in the sense that for every pair of times (s, t) with $s \leq t$, the two conditions $(X_u)_{u=1}^t = (\widetilde{X}_u)_{u=1}^t$ and $V_t(X) \leq V_t(\widetilde{X})$ together imply $V_s(X) \leq V_s(\widetilde{X})$.

Proof of Proposition 2. (i) The properties follow immediately from (18). (ii) In proving time consistency it is sufficient to take $s = t - 1$. By (15),

$$V_{t-1}(X) = W_{t-1}(X_t + V_t(X))$$
$$= W_{t-1}(\widetilde{X}_t + V_t(X))$$
$$\leq W_{t-1}(\widetilde{X}_t + V_t(\widetilde{X}))$$
$$= V_{t-1}(\widetilde{X}).$$

$\square$

The dynamic version of Value-at-Risk presented in the example below is an example of a dynamic monetary risk measure $(R_t)_{t=0}^{T-1}$ with $R_t : L^0(\mathcal{F}_{t+1}) \to L^0(\mathcal{F}_t)$. In Section 3.3 further examples of dynamic monetary risk measures and utility functions are presented and their properties are investigated for use in Section 4 together with specific models for the liability cash flows.

Example 1. *In the static or one-period setting, Value-at-Risk at time 0 at level $u \in (0,1)$ of a value $Z \in L^0(\mathcal{F}_1)$ is defined as*

$$\begin{aligned}
\mathrm{VaR}_u(Z) &:= \min\{m \in \mathbb{R} : \mathbb{P}(m + Z < 0) \leq u\} \\
&= \min\{m \in \mathbb{R} : \mathbb{P}(-Z \leq m) \geq 1 - u\} \\
&= \min\{m \in \mathbb{R} : Q_{-Z}((-\infty, m]) \geq 1 - u)\} \\
&=: F_{-Z}^{-1}(1 - u),
\end{aligned}$$

where Q_{-Z} denotes the distribution of $-Z$, and $F_{-Z}(m) = Q_{-Z}((-\infty, m])$. The natural dynamic version of Value-at-Risk at time t at level u of a value $Z \in L^0(\mathcal{F}_{t+1})$ is

$$\mathrm{VaR}_{t,u}(Z) := \mathrm{ess}\inf\{m \in L^0(\mathcal{F}_t) : \mathbb{P}(m + Z < 0 \mid \mathcal{F}_t) \leq u\},$$

where "ess inf" denotes the greatest lower bound of a family of random variables (with respect to $\mathbb{P}$-almost sure inequality). Alternatively, we may define $\mathrm{VaR}_{t,u}(Z)$ in terms of a conditional distribution $Q_{t,-Z}$ of $-Z$ with respect to $\mathcal{F}_t$: for each $\omega \in \Omega$, $Q_{t,-Z}(\omega, \cdot)$ is a probability measure on the Borel subsets of $\mathbb{R}$, and for each Borel set $A \subset \mathbb{R}$, $Q_{t,-Z}(\cdot, A)$ is a version of $\mathbb{P}(-Z \in A \mid \mathcal{F}_t)$. Define $Z' \in L^0(\mathcal{F}_t)$ by

$$Z'(\omega) := \min\{m \in \mathbb{R} : Q_{t,-Z}(\omega, (-\infty, m]) \geq 1 - u\} =: F_{t,-Z}^{-1}(\omega, 1 - u)$$

and notice that $\mathrm{VaR}_{t,u}(Z) = Z'$ $\mathbb{P}$-almost surely. Notice that $(\mathrm{VaR}_{t,u})_{t=0}^{T-1}$ satisfies the properties in Definition 1 for $p = 0$.

3.1. **Model-invariant bounds.** Consider a dynamic risk measure $(R_t)_{t=0}^{T-1}$ and a dynamic monetary utility function $(U_t)_{t=0}^{T-1}$, and $(W_t)_{t=0}^{T-1}$ given by (13). Consider also an $\mathbb{F}$-adapted cash flow $(X_t)_{t=1}^T$ with $X_t \in L^p(\mathcal{F}_t)$ for every t. With $Y_{t+1} := X_{t+1} + V_{t+1}(X)$, combining (8) and (9) yields

$$\begin{aligned}
V_t(X) &= R_t(-Y_{t+1}) - \frac{1}{1 + \eta_t} U_t((R_t(-Y_{t+1}) - Y_{t+1})_+) \\
&\leq R_t(-Y_{t+1}) \\
&= R_t(-X_{t+1} - V_{t+1}(X)).
\end{aligned}$$

Repeated application of this inequality together with (4) and (5) gives the upper bound

$$\begin{aligned}
V_t(X) &\leq R_t(-X_{t+1} - R_{t+1}(-X_{t+2} - \cdots - R_{T-1}(-X_T) \ldots)) \\
&= R_t(-R_{t+1}(\cdots - R_{T-1}(-X_{t+1} - \cdots - X_T) \ldots)).
\end{aligned}$$

Further assumptions clearly enable sharper bounds. Suppose that $p \geq 1$ and take, for every t, U_t to be the conditional expectation given $\mathcal{F}_t$. It follows from Jensen's inequality for conditional expectations that the conditional

expectation is well defined as a mapping $L^p(\mathcal{F}_u) \to L^p(\mathcal{F}_t)$ for $u > t$. In particular,

$$V_t(X) = R_t(-Y_{t+1}) - \frac{1}{1+\eta_t}\mathbb{E}\left[\left(R_t(-Y_{t+1}) - Y_{t+1}\right)_+ \mid \mathcal{F}_t\right]$$

$$\leq R_t(-Y_{t+1}) - \frac{1}{1+\eta_t}\mathbb{E}\left[R_t(-Y_{t+1}) - Y_{t+1} \mid \mathcal{F}_t\right]$$

(19) $$= \frac{1}{1+\eta_t}\left(\eta_t R_t(-Y_{t+1}) + \mathbb{E}\left[X_{t+1} \mid \mathcal{F}_t\right] + \mathbb{E}\left[V_{t+1}(X) \mid \mathcal{F}_t\right]\right).$$

Applying this inequality repeatedly together with the tower property of conditional expectation yields

$$V_t(X) \leq \sum_{s=t}^{T-1}\mathbb{E}\left[\frac{\eta_s R_s(-Y_{s+1})}{\prod_{u=t}^s(1+\eta_u)} \mid \mathcal{F}_t\right] + \sum_{s=t}^{T-1}\mathbb{E}\left[\frac{X_{s+1}}{\prod_{u=t}^s(1+\eta_u)} \mid \mathcal{F}_t\right].$$

Notice that if $\eta_t = \eta_0$ for all t, then

$$V_0(X) \leq \eta_0\sum_{t=0}^{T-1}\frac{\mathbb{E}\left[R_t(-Y_{t+1})\right]}{(1+\eta_0)^{t+1}} + \sum_{t=1}^T\frac{\mathbb{E}\left[X_t\right]}{(1+\eta_0)^t}.$$

If the static replicating portfolio is chosen at time 0 such that $\mathbb{E}\left[X^s\right] = \mathbb{E}\left[X^o\right]$, then the residual cash flow $X := X^o - \mathbb{E}\left[X^o\right]$ has zero mean so the second sum in the above upper bound vanishes, i.e.

(20) $$V_0(X) \leq \eta_0\sum_{t=0}^{T-1}\frac{\mathbb{E}\left[R_t(-Y_{t+1})\right]}{(1+\eta_0)^{t+1}}.$$

Notice that the upper bound for $V_t(X)$ in (19) can be rewritten as

$$V_t(X) \leq \eta_t(R_t(-Y_{t+1}) - V_t(X)) + \mathbb{E}\left[X_{t+1} \mid \mathcal{F}_t\right] + \mathbb{E}\left[V_{t+1}(X) \mid \mathcal{F}_t\right].$$

Repeated application of this inequality and use of the tower property of conditional expectation yields

$$V_0(X) \leq \sum_{t=0}^{T-1}\mathbb{E}\left[\eta_t\left(R_t(-Y_{t+1}) - V_t(X)\right)\right] + \sum_{t=1}^T\mathbb{E}\left[X_t\right].$$

Hence, if further $\eta_t = \eta_0$ for all t and $\mathbb{E}\left[X\right] = 0$, then

(21) $$V_0(X) \leq \eta_0\sum_{t=0}^{T-1}\mathbb{E}\left[R_t(-Y_{t+1}) - V_t(X)\right].$$

Notice the difference between the two upper bounds (20) and (21): the former is formulated in terms of expected future capital requirement whereas the latter is formulated in terms of expected future buffer capital provided by capital providers.

We end the discussion of model-invariant bounds for the cost-of-capital margin with a comment on the Solvency II risk margin. In [9, Article 37] it is stated that the risk margin should be computed as

$$\text{CoC}\sum_{t\geq 0}\frac{\text{SCR}(t)}{(1+r(t+1))^{t+1}},$$

where $\text{CoC} := 0.06$, $r(t+1)$ denotes the basic risk-free interest rate for the maturity of $t + 1$ years, and $\text{SCR}(t)$ denotes the Solvency Capital Requirement after t years. In our setting, $\text{SCR}(t) = 0$ for $t \geq T$ since there is no liability cash flow beyond that time. One may criticize several aspects of the Solvency II risk margin. First, for $t \geq 1$, $\text{SCR}(t)$ is a random variable as seen from time 0. Secondly, $\text{SCR}(t)$ does not take capital costs into account, and the discounting in the computation of $\text{SCR}(t)$ and in the computation of the risk margin are not conceptually consistent.

The upper bound in (21) is somewhat similar to the formula for the Solvency II risk margin. For $p \geq 1$, take $(R_t)_{t=0}^{T-1}$ to be a dynamic risk measure in the sense of Definition 1, and take the U_t to be conditional expectations $\mathbb{E}[\cdot \mid \mathcal{F}_t]$. If we define the SCR-like quantity

$$\widetilde{\text{SCR}}_t(X^o) := R_t\Big(\sum_{u=t+1}^{T} \mathbb{E}[X_u^o \mid \mathcal{F}_t] + V_t(X^o - \mathbb{E}[X^o \mid \mathcal{F}_t])$$
$$- X_{t+1}^o - \sum_{u=t+2}^{T} \mathbb{E}[X_u^o \mid \mathcal{F}_{t+1}] - V_{t+1}(X^o - \mathbb{E}[X^o \mid \mathcal{F}_{t+1}])\Big),$$

then, using the the second of the three properties in Proposition 2 (i) for V_t and V_{t+1}, it can be shown that $\widetilde{\text{SCR}}_t(X^o) = R_t(-Y_{t+1}) - V_t(X)$, where, as before, $X := X^o - \mathbb{E}[X^o \mid \mathcal{F}_0]$ and $Y_{t+1} := X_{t+1} + V_{t+1}(X)$. Indeed,

$$V_t(X^o - \mathbb{E}[X^o \mid \mathcal{F}_t]) - V_{t+1}(X^o - \mathbb{E}[X^o \mid \mathcal{F}_{t+1}])$$
$$= V_t(X) - V_{t+1}(X) + \mathbb{E}[X_{t+1}^o \mid \mathcal{F}_0] - \mathbb{E}[X_{t+1}^o \mid \mathcal{F}_t]$$
$$+ \sum_{u=t+2}^{T} \big(\mathbb{E}[X_u^o \mid \mathcal{F}_{t+1}] - \mathbb{E}[X_u^o \mid \mathcal{F}_t]\big),$$

which yields

$$\widetilde{\text{SCR}}_t(X^o) = R_t\big(\mathbb{E}[X_{t+1}^o \mid \mathcal{F}_0] - \mathbb{E}[X_{t+1}^o \mid \mathcal{F}_{t+1}] + V_t(X) - V_{t+1}(X)\big)$$
$$= R_t(-Y_{t+1}) - V_t(X).$$

In particular, (21) can be rephrased as

$$V_0(X) \leq \eta_0 \sum_{t=0}^{T-1} \mathbb{E}\left[\widetilde{\text{SCR}}_t(X^o)\right].$$

For an in-depth comparison between a conceptually consistent notion of cost-of-capital margin and the Solvency II risk margin, see Section 5 in [11].

3.2. **Dynamic monetary utility functions.** The notions of risk-adjusted values in [3] and conditional and dynamic monetary utility functions in [4] are closely connected to the cost-of-capital margin considered here. In [3] and [4], cumulative cash flows and value processes are considered whereas we consider incremental cash flows and liability processes corresponding to liability values for future cash flows. We will now proceed to establish the connection between $V_t(\cdot)$ (and $W_t(\cdot)$) and the work of [3] and [4].

For $p \in [0, \infty]$, let $\mathcal{R}_{1,T}^p$ denote the space of all $\mathbb{F}$-adapted stochastic processes $(Y_t)_{t=1}^{T}$ with $Y_t \in L^p(\mathcal{F}_t)$ for every t. For $1 \leq u \leq v \leq T$, define

the projection $\pi_{u,v} : \mathcal{R}^p_{1,T} \to \mathcal{R}^p_{1,T}$ by

$$\pi_{u,v}(Y)_t := 1_{u \le t} Y_{t \wedge v}, \quad t \in \{1, \dots, T\},$$

and $\mathcal{R}^p_{u,v} := \pi_{u,v} \mathcal{R}^p_{1,T}$. Let $(W_t)^{T-1}_{t=0}$ be as in Proposition 1. For all t, let $\varphi_{t,T} : \mathcal{R}^p_{t,T} \to L^p(\mathcal{F}_t)$ be given by

(22) $$\varphi_{t,T}(Y) := -W_t \circ \cdots \circ W_{T-1}(-Y_T).$$

In [3] and [4] the elements in $\mathcal{R}^p_{1,T}$ are interpreted as cumulative rather than incremental cash flows. For an incremental cash flow $X \in \mathcal{R}^p_{1,T}$, let $X^c \in \mathcal{R}^p_{1,T}$ be the cumulative cash flow given by $X^c_t := \sum^t_{s=1} X_s$, and notice that

$$\varphi_{t,T}(X^c) = X^c_t - W_t \circ \cdots \circ W_{T-1}(-(X^c_T - X^c_t))$$
$$= \sum^t_{s=1} X_s - V_t(-X),$$

and similarly, $V_t(X) = -X^c_t - \varphi_{t,T}(-X^c)$.

We now verify that $\varphi_{t,T}$ in (22) is a conditional monetary utility function in the sense of Definition 3.1 in [4], excluding the concavity axiom, and that $(\varphi_{t,T})^T_{t=1}$ is time-consistent in the sense of Definition 4.2 in [4].

Proposition 3. *The mappings $\varphi_{t,T}$ in (22) are conditional monetary utility functions in the sense*

$$\varphi_{t,T}(0) = 0,$$

$$\varphi_{t,T}(Y) \le \varphi_{t,T}(\widetilde{Y}) \text{ for all } Y, \widetilde{Y} \in \mathcal{R}^p_{t,T} \text{ such that } Y \le \widetilde{Y},$$

$$\varphi_{t,T}(Y + m1_{[t,T]}) = \varphi_{t,T}(Y) + m \text{ for all } Y \in \mathcal{R}^p_{t,T} \text{ and } m \in L^p(\mathcal{F}_t),$$

and $(\varphi_{t,T})^T_{t=1}$ is time-consistent in the sense

(23) $$\varphi_{t,T}(Y) = \varphi_{t,T}(Y1_{[t,u)} + \varphi_{u,T}(Y)1_{[u,T]})$$

for every $t \le u \le T$ and all $X \in \mathcal{R}^p_{t,T}$.

Proof of Proposition 3. Since $W_s(0) = 0$ for $s = 0, \dots, T-1$,

$$\varphi_{t,T}(0) = -W_t \circ \cdots \circ W_{T-1}(-0) = 0.$$

Due to the fact that $W_s(0) = 0$ for $s = 0, \dots, T-1$. For $Y, \widetilde{Y} \in \mathcal{R}^p_{t,T}$ such that $Y \le \widetilde{Y}$,

$$\varphi_{t,T}(\widetilde{Y}) - \varphi_{t,T}(Y) = -W_t \circ \cdots \circ W_{T-1}(-\widetilde{Y}_T) + W_t \circ \cdots \circ W_{T-1}(-Y_T)$$

Noting that $-Y_T \ge -\widetilde{Y}_T$ and using, repeatedly, the monotonicity property of W_s for $s = 0, \dots, T-1$ we arrive at $\varphi_{t,T}(\widetilde{Y}) - \varphi_{t,T}(Y) \ge 0$. Finally we note that

$$\varphi_{t,T}(Y + m1_{[t,T]}) = -W_t \circ \cdots \circ W_{T-1}(-(Y_T + m))$$
$$= m - W_t \circ \cdots \circ W_{T-1}(-Y_T)$$
$$= \varphi_{t,T}(Y) + m,$$

where the second equality follows from the translation invariance of W_s for $s = t, \ldots, T-1$. Time consistency in the sense of (23) follows almost directly from the definition.

$$
\begin{aligned}
\varphi_{t,T}&(Y1_{[t,u)} + \varphi_{u,T}(Y)1_{[u,T]}) \\
&= -W_t \circ \cdots \circ W_{T-1}(-(Y1_{[t,u)} + \varphi_{u,T}(Y)1_{[u,T]})_T) \\
&= -W_t \circ \cdots \circ W_{T-1}(-\varphi_{u,T}(Y)) \\
&= -W_t \circ \cdots \circ W_{u-1}(-\varphi_{u,T}(Y)),
\end{aligned}
$$

where the last equality is due to the translation invariance of W_s for $s = t, \ldots, T-1$ combined with the fact that $-\varphi_{u,T}(Y)$ is $\mathcal{F}_s$-measurable for $s \geq u$ and noting that $W_s(0) = 0$ for $s = u, \ldots, T-1$. From the definition (22) of $\varphi_{u,T}$ we get

$$
\begin{aligned}
-&W_t \circ \cdots \circ W_{u-1}(-(-W_u \circ \cdots \circ W_{T-1}(-Y_T))) \\
&= -W_t \circ \cdots \circ W_{T-1}(-Y_T) \\
&= \varphi_{t,T}(Y).
\end{aligned}
$$

$\square$

3.3. **Risk measures and utility functions based on conditional quantiles.** For $Y \in L^p(\mathcal{F}_{t+1})$, write $Q_{t,Y}$ for its conditional distribution given $\mathcal{F}_t$: $Q_{t,Y}(\omega, \cdot)$ is a probability measure on the Borel subsets of $\mathbb{R}$, and $Q_{t,Y}(\cdot, A)$ is a version of $\mathbb{P}(Y \in A \mid \mathcal{F}_t)$. We may write the conditional distribution and quantile functions of Y given $\mathcal{F}_t$, respectively, as

$$
\begin{aligned}
F_{t,Y}(\omega, y) &:= Q_{t,Y}(\omega, (-\infty, y]), \\
F_{t,Y}^{-1}(\omega, u) &:= \min\{y \in \mathbb{R} : F_{t,Y}(\omega, y) \geq u\}.
\end{aligned}
$$

For a probability measures M^R and M^U on $(0, 1)$, define, for $Y \in L^p(\mathcal{F}_{t+1})$,

$$
(24) \qquad R_t(Y) := \int_0^1 F_{t,-Y}^{-1}(u) dM^R(u),
$$

$$
(25) \qquad U_t(Y) := \int_0^1 F_{t,Y}^{-1}(u) dM^U(u).
$$

Remark 4. *Conditional monetary risk measures of the form* (24) *include risk measures such as Value-at-Risk and Expected Shortfall that are prescribed by current solvency regulation for capital requirements:* $\mathrm{VaR}_{t,0.005}(Y) = F_{t,-Y}^{-1}(0.995)$ *is obtained by taking M^R to be the degenerated probability distribution corresponding to a unit point mass at 0.995, and*

$$
\mathrm{ES}_{t,0.01}(Y) := \frac{1}{0.01} \int_0^{0.01} \mathrm{VaR}_{t,u}(Y) du
$$

is obtained by taking M^R to be the uniform distribution on $(0.99, 1)$.

Conditional monetary utility functions of the form (25) *are utility functions in the sense of Yaari's Dual Theory of Choice Under Risk, see* [26]. *Alternatively, when applied to a random variable representing a loss, they are -1 times a distortion risk measures in the sense of Wang, see* [21] *and* [22]. *This is described below.*

A distortion function is a nondecreasing function $g : [0,1] \to [0,1]$ satisfying $g(0) = 0$ and $g(1) = 1$. Given a distortion function g, the g-distorted expectation of Y, with survival function $\overline{F}_Y$, is given by

$$(26) \qquad \rho_g(Y) := -\int_{-\infty}^{0} \left(1 - g\left(\overline{F}_Y(x)\right)\right)dx + \int_{0}^{\infty} g\left(\overline{F}_Y(x)\right)dx$$

$$(27) \qquad\qquad = \int_{0}^{1} F_Y^{-1}((1-u)+)dg(u),$$

where $F_Y^{-1}((1-u)+) := \lim_{v \downarrow 1-u} F_Y^{-1}(v)$. The equality (27) follows from Theorem 4 in [7]. Notice that the first integral in (26) vanishes if only nonnegative random variables are considered. Finally, notice that

$$
\begin{aligned}
U_t(Y) &:= \int_{0}^{1} F_{t,Y}^{-1}(u)dM^U(u) \\
&= -\int_{0}^{1} F_{t,-Y}^{-1}((1-u)+)dM^U(u) \\
&= -\rho_{g,t}(-Y),
\end{aligned}
$$

where g is the distribution function of M^U and $\rho_{g,t}(-Y)$ is the $\mathcal{F}_t$-measurable g-distorted expectation of $-Y$ obtained by replacing F_Y by $F_{t,Y}$ in (26). If g is continuous, then $U_t(Y) = -\rho_{g,t}(-Y) = \rho_{g,t}(Y)$.

Proposition 4. *Suppose there exist $u_0 \in (0, 1/2)$ and $\overline{m} \in (0, \infty)$ such that, for $k = R, U$,*

$$\max\left(M^k((u,v)), M^k((1-v, 1-u))\right) \le \overline{m}(v-u) \quad \text{for all } 0 < u < v < u_0.$$

Fix $p \in [1, \infty]$.
(i) R_t in (24) and U_t in (25) are well-defined as mappings from $L^p(\mathcal{F}_{t+1})$ to $L^p(\mathcal{F}_t)$ and satisfy (4)-(6) and (7)-(9), respectively.
(ii) If $Y \in L^p(\mathcal{F}_{t+1})$ and, for any Borel set $A \subset \mathbb{R}$

$$\mathbb{P}(Y \in A \mid \mathcal{F}_t) = \mathbb{P}(Y^{(1)} + Y^{(2)}Y^{(3)} \in A \mid \mathcal{F}_t),$$

where $Y^{(1)} \in L^p(\mathcal{F}_t)$, $0 < Y^{(2)} \in L^0(\mathcal{F}_0)$, and $Y^{(3)} \in L^p(\mathcal{F}_{t+1})$ is independent of $\mathcal{F}_t$, then

$$R_t(Y) = -Y^{(1)} + Y^{(2)}R_t(Y^{(3)}),$$
$$U_t(Y) = Y^{(1)} + Y^{(2)}U_t(Y^{(3)}),$$

where $R_t(Y^{(3)}), U_t(Y^{(3)}) \in L^0(\mathcal{F}_0)$. Moreover, $R_t(Y^{(3)}) = R_0(\widetilde{Y}^{(3)})$ and $U_t(Y^{(3)}) = U_0(\widetilde{Y}^{(3)})$ for $\widetilde{Y}^{(3)} \in L^p(\mathcal{F}_1)$ with $Y^{(3)}$ and $\widetilde{Y}^{(3)}$ equally distributed. For W_t in (13),

$$W_t(Y) = Y^{(1)} + Y^{(2)}W_t(Y^{(3)}),$$

and if $\eta_t \in L_+^0(\mathcal{F}_0)$, then $W_t(Y^{(3)}) \in L^0(\mathcal{F}_0)$. Further, if $\eta_t = \eta_0$, then $W_t(Y^{(3)}) = W_0(\widetilde{Y}^{(3)})$ for $\widetilde{Y}^{(3)} \in L^p(\mathcal{F}_1)$ with $Y^{(3)}$ and $\widetilde{Y}^{(3)}$ equally distributed.

Proof of Proposition 4. (i) It is sufficient to prove the statement (i) for R_t and U_t only for R_t since the same proof, with minor modifications, applies to U_t. For $p = \infty$ the statement holds without the requirement on M^R.

We now consider $p \in [1, \infty)$. The conditional quantile has the monotonicity property

$$F_{t,Y}^{-1}(\omega, u) \leq F_{t,\widetilde{Y}}^{-1}(\omega, u) \quad \text{if } Y \leq \widetilde{Y}$$

and the properties $F_{t,Y}^{-1}(\omega, u) \leq 0$ if $Y \leq 0$ and $F_{t,Y}^{-1}(\omega, u) \geq 0$ if $Y \geq 0$. In particular, $R_t(|Y|) \leq R_t(Y) \leq R_t(-|Y|)$. We show that $R_t(|Y|), R_t(-|Y|) \in L^p(\mathcal{F}_t)$.

$$
\begin{aligned}
\mathbb{E}\left[R_t(-|Y|)^p\right] &= \mathbb{E}\left[\left(\int_0^1 F_{t,|Y|}^{-1}(u) dM^R(u)\right)^p\right] \\
&\leq \mathbb{E}\left[\int_0^1 \left(F_{t,|Y|}^{-1}(u)\right)^p dM^R(u)\right] \\
&= \mathbb{E}\left[\int_0^{1-u_0} F_{t,|Y|^p}^{-1}(u) dM^R(u)\right] + \mathbb{E}\left[\int_{1-u_0}^1 F_{t,|Y|^p}^{-1}(u) dM^R(u)\right] \\
&=: E_1 + E_2,
\end{aligned}
$$

where the first inequality above is an application of Jensen's inequality, and the second equality follows from the fact that $g(F_Z^{-1}(u)) = F_{g(Z)}^{-1}(u)$ for increasing functions g. Moreover,

$$
\begin{aligned}
E_1 &\leq \mathbb{E}\left[F_{t,|Y|^p}^{-1}(1-u_0) \int_0^{1-u_0} dM^R(u)\right] \leq \mathbb{E}\left[F_{t,|Y|^p}^{-1}(1-u_0)\right], \\
E_2 &\leq \overline{m}\mathbb{E}\left[\int_{1-u_0}^1 F_{t,|Y|^p}^{-1}(u) du\right] \leq \overline{m}\mathbb{E}\left[\int_0^1 F_{t,|Y|^p}^{-1}(u) du\right].
\end{aligned}
$$

Since

$$\mathbb{E}\left[\int_0^1 F_{t,|Y|^p}^{-1}(u) du\right] = \mathbb{E}\left[\mathbb{E}\left[|Y|^p \mid \mathcal{F}_t\right]\right] = \mathbb{E}\left[|Y|^p\right] < \infty$$

and

$$\mathbb{E}\left[\int_0^1 F_{t,|Y|^p}^{-1}(u) du\right] \geq \mathbb{E}\left[\int_{1-u_0}^1 F_{t,|Y|^p}^{-1}(u) du\right] \geq u_0 \mathbb{E}\left[F_{t,|Y|^p}^{-1}(1-u_0)\right],$$

$E_1, E_2 < \infty$, from which $\mathbb{E}\left[|R_t(-|Y|)|^p\right] = \mathbb{E}\left[R_t(-|Y|)^p\right] < \infty$ follows. The argument for showing $\mathbb{E}\left[|R_t(|Y|)|^p\right] < \infty$ is completely analogous upon writing

$$
\begin{aligned}
\mathbb{E}\left[(-R_t(|Y|))^p\right] &= \mathbb{E}\left[\left(\int_0^1 (-F_{t,-|Y|}^{-1}(u)) dM^R(u)\right)^p\right] \\
&= \mathbb{E}\left[\left(\int_0^1 F_{t,|Y|}^{-1}((1-u)+) dM^R(u)\right)^p\right],
\end{aligned}
$$

where $F_{t,|Y|}^{-1}(\omega, (1-u)+) := \lim_{v \downarrow 1-u} F_{t,|Y|}^{-1}(\omega, v)$.

Set $Y := Y^{(1)} + Y^{(2)}Y^{(3)}$, where $Y^{(1)} \in L^p(\mathcal{F}_t)$, $0 < Y^{(2)} \in L^p(\mathcal{F}_t)$, and $Y^{(3)} \in L^p(\mathcal{F}_{t+1})$. Then

$$
\begin{aligned}
F_{t,Y}(\omega, y) &:= Q_{t,Y}(\omega, (-\infty, y]) \\
&= Q_{t,Y^{(3)}}(\omega, (-\infty, (y - Y^{(1)}(\omega))/Y^{(2)}(\omega)]).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
F_{t,Y}^{-1}(\omega, u) &:= \min\{y \in \mathbb{R} : Q_{t,Y}(\omega, (-\infty, y]) \geq u\} \\
&= \min\{y \in \mathbb{R} : Q_{t,Y^{(3)}}(\omega, (-\infty, (y - Y^{(1)}(\omega))/Y^{(2)}(\omega)]) \geq u\} \\
&= Y^{(1)}(\omega) + Y^{(2)}(\omega) \min\{y \in \mathbb{R} : Q_{t,Y^{(3)}}(\omega, (-\infty, y]) \geq u\} \\
&= Y^{(1)}(\omega) + Y^{(2)}(\omega) F_{t,Y^{(3)}}^{-1}(\omega, u).
\end{aligned}
$$

Similarly, $F_{t,-Y}^{-1}(\omega, u) = -Y^{(1)}(\omega) + Y^{(2)}(\omega) F_{t,-Y^{(3)}}^{-1}(\omega, u)$. It now follows from the definitions of R_t and U_t in (24) and (25) that the properties (4)-(6) and (7)-(9) hold. The proof of statement (i) is complete.

(ii) Under the stronger assumption that $Y^{(1)} \in L^p(\mathcal{F}_t)$, $0 < Y^{(2)} \in L^0(\mathcal{F}_0)$, and $Y^{(3)} \in L^p(\mathcal{F}_{t+1})$ is independent of $\mathcal{F}_t$,

$$
\begin{aligned}
F_{t,Y}^{-1}(\omega, u) &:= Y^{(1)}(\omega) + Y^{(2)}(\omega) F_{t,Y^{(3)}}^{-1}(\omega, u) \\
&= Y^{(1)}(\omega) + Y^{(2)} F_{Y^{(3)}}^{-1}(u)
\end{aligned}
$$

since $Y^{(2)}$ is a constant and $Y^{(3)}$ does not depend on $\mathcal{F}_t$. Similarly,

$$
F_{t,-Y}^{-1}(\omega, u) = -Y^{(1)}(\omega) + Y^{(2)} F_{-Y^{(3)}}^{-1}(u).
$$

It follows from the definitions of R_t, U_t and W_t in (24), (25) and (13), that $W_t(Y) = Y^{(1)} + Y^{(2)} W_t(Y^{(3)})$, and similarly for $R_t(Y)$ and $U_t(Y)$, where $W_t(Y^{(3)})$ is a constant if $\eta_t \in L_+^0(\mathcal{F}_0)$. $\qquad\square$

Remark 5. *Notice that the condition on M^R in Proposition 4 holds e.g. if M^R has support in $(0,1)$ bounded away from 0 and 1 (the case for a conditional version of Value-at-Risk) or if M^R has a bounded density (the case for a conditional version of Expected Shortfall).*

Example 2. *Here we derive an expression for $W_t(Y_{t+1})$ for R_t and U_t of the form (24) and (24), respectively, under the assumption that the distribution function $x \mapsto M^U(0,x]$ is continuous. Using well-known properties of quantile functions, we may write*

$$
\begin{aligned}
&\int_0^1 F_{t,(R_t(-Y_{t+1})-Y_{t+1})_+}^{-1}(u)\,dM^U(u) \\
&= -\int_0^1 F_{t,-(R_t(-Y_{t+1})-Y_{t+1})_+}^{-1}(1-u)\,dM^U(u) \\
&= \int_0^1 (R_t(-Y_{t+1}) - F_{t,Y_{t+1}}^{-1}(1-u))_+\,dM^U(u) \\
&= R_t(-Y_{t+1})\int_{1-\gamma_t}^1 dM^U(u) - \int_{1-\gamma_t}^1 F_{t,Y_{t+1}}^{-1}(1-u)\,dM^U(u),
\end{aligned}
$$

where $\gamma_t := \mathbb{P}(Y_{t+1} \leq R_t(-Y_{t+1}) \mid \mathcal{F}_t)$. Hence,

$$
\begin{aligned}
W_t(Y_{t+1}) = {}& \Big(1 - \frac{M^U(1-\gamma_t, 1)}{1+\eta_t}\Big)\int_0^1 F_{t,Y_{t+1}}^{-1}(u)\,dM^R(u) \\
& + \frac{1}{1+\eta_t}\int_{1-\gamma_t}^1 F_{t,Y_{t+1}}^{-1}(1-u)\,dM^U(u).
\end{aligned}
$$

In particular,

$$W_t(Y_{t+1}) = \int_0^1 F_{t,Y_{t+1}}^{-1}(u) dM_t^W(u),$$

where M_t^W is an $\mathcal{F}_t$-measurable probability distribution on $(0,1)$ given by

$$M_t^W(A) := \Big(1 - \frac{M^U(1-\gamma_t, 1)}{1+\eta_t}\Big) M^R(A) + \frac{\widetilde{M}^U(A \cap (0, \gamma_t))}{1+\eta_t},$$

where $\widetilde{M}^U := M^U \circ h^{-1}$ for $h(u) := 1 - u$ and A is a Borel subset of $(0,1)$.

If $R_t = \mathrm{VaR}_{t,p}$, $\eta_t = \eta_0$ and $F_{t,Y_{t+1}}$ is continuous, then $\gamma_t = 1 - p$ and $M_t^W = M_0^W$. If further $U_t(\cdot) = \mathbb{E}[\cdot \mid \mathcal{F}_t]$, then

$$W_t(Y_{t+1}) = \frac{1}{1+\eta_0}\Big((\eta_0 + p)\mathrm{VaR}_{t,p}(-Y_{t+1}) - p\,\mathrm{ES}_{t,p}(-Y_{t+1})$$
$$+ \mathbb{E}[Y_{t+1} \mid \mathcal{F}_t]\Big).$$

If $dM^R(u) = m^R(u)du$ and $dM^U(u) = m^U(u)du$ for monotone integrable functions m^R and m^U with m^R nondecreasing and m^U nonincreasing, then M_t^W has a density

$$m_t^W(u) = \Big(1 - \frac{M^U(1-\gamma_t, 1)}{1+\eta_t}\Big) m^R(u) + \frac{1}{1+\eta_t} m^U(1-u) 1_{[0,\gamma_t]}(u).$$

The density m_t^W is not monotone. Similarly to the argument in the proof of Theorem 4.1 in [1], it follows that W_t does not have the subadditivity property $W_t(Y_{t+1}+\widetilde{Y}_{t+1}) \leq W_t(Y_{t+1})+W_t(\widetilde{Y}_{t+1})$. However, see Proposition 9 below, we may ensure subadditivity of the cost-of-capital margin by imposing restrictions on the stochastic model for the liability cash flow.

4. Cost-of-capital margin for specific models

In order to obtain stronger results we need to impose further assumptions. We will therefore assume conditional risk measures R_t and conditional utility functions U_t of the kind presented in Section 3.3. Moreover, we will consider flexible models that, when combined with those conditional risk measures and utility functions provide e.g. explicit formulas for the cost-of-capital margin. More specifically, in Section 4.1 we consider residual cash flows of an autoregressive form, and in 4.2 we consider a class of Gaussian models for the residual cash flows.

4.1. **Autoregressive cash flows.** Residual cash flows that are given by an autoregressive process of order one are particularly well suited for explicit computation of the cost-of-capital margin when the dynamic monetary risk measures and dynamic monetary utility functions are of the type presented in Section 3.3. The autoregressive processes include residual cash flows with independent components as a special case.

Proposition 5. *Fix $p \in [1, \infty]$ and let W_t be given by (13) with R_t and U_t in (24) and (25), respectively, satisfying the condition in Proposition 4. Let $(Z_t)_{t=1}^T$ be an $\mathbb{F}$-adapted sequence of random variables such that, for each t,*

$Z_{t+1} \in L^p(\mathcal{F}_{t+1})$ is independent of $\mathcal{F}_t$. Let $(\alpha_t)_{t=1}^T$ be a nonrandom sequence of real numbers. Let

$$X_0 := 0, \quad X_{t+1} := \alpha_{t+1}X_t + Z_{t+1}, \quad t = 0, \dots, T-1,$$

and set

$$\beta_T := 1, \quad \beta_t := 1 + \beta_{t+1}\alpha_{t+1}, \quad t \in \{1, \dots, T-1\},$$
$$\delta_T := 0, \quad \delta_t := \delta_{t+1} + |\beta_{t+1}|W_t(\operatorname{sign}(\beta_{t+1})Z_{t+1}), \quad t \in \{0, \dots, T-1\}.$$

Then, for $t = 0, \dots, T-1$, $\delta_t \in L^0(\mathcal{F}_0)$ and

$$(28) \qquad\qquad V_t(X) = \delta_t + \beta_{t+1}\alpha_{t+1}X_t \in L^p(\mathcal{F}_t).$$

In particular,

$$V_0(X) = \sum_{t=0}^{T-1} |\beta_{t+1}|W_t(\operatorname{sign}(\beta_{t+1})Z_{t+1}).$$

Proof of Proposition 5. The statement is proved by induction. First,

$$V_{T-1}(X) = W_{T-1}(X_T) = \alpha_T X_{T-1} + W_{T-1}(Z_T) = \delta_{T-1} + \beta_T \alpha_T X_{T-1},$$

i.e. (28) holds for $t = T-1$. The recursion step: take $t \in \{1, \dots, T-1\}$ and suppose that $V_t(X)$ is given by (28). Then

$$X_t + V_t(X) = \delta_t + \beta_t X_t = \delta_t + \beta_t \alpha_t X_{t-1} + \beta_t Z_t.$$

From Proposition 4 it follows that $W_{t-1}(\beta_t Z_t)$ is a constant and $X_t + V_t(X) \in L^p(\mathcal{F}_t)$. Moreover,

$$\begin{aligned}
V_{t-1}(X) &= W_{t-1}(X_t + V_t(X)) \\
&= \delta_t + \beta_t \alpha_t X_{t-1} + W_{t-1}(\beta_t Z_t) \\
&= \delta_t + \beta_t \alpha_t X_{t-1} + |\beta_t|W_{t-1}(\operatorname{sign}(\beta_t)Z_t) \\
&= \delta_{t-1} + \beta_t \alpha_t X_{t-1}.
\end{aligned}$$

In particular, $V_{t-1}(X) = W_{t-1}(X_t + V_t(X)) \in L^p(\mathcal{F}_{t-1})$. We conclude that (28) holds for $t = 0, \dots, T-1$. $\qquad\square$

Remark 6. *Notice that in the special case where $(X_t)_{t=1}^T$ has independent components, corresponding to $\alpha_t = 0$ for all t, $V_t(X) = \sum_{s=t}^{T-1} W_s(X_{s+1})$. In particular, $V_t(X) \in L^0(\mathcal{F}_0)$ if $\eta_t \in L_+^0(\mathcal{F}_0)$ for all t. If $\alpha_t = \alpha \in (-1, 1)$ for all t, then*

$$0 < \beta_t = \sum_{j=0}^{T-t} \alpha^j = \frac{1 - \alpha^{T-t+1}}{1 - \alpha}.$$

If further $W_t(Z_{t+1}) = W_0(Z_1)$ for all t, then $V_0(X) = f(\alpha)W_0(Z_1)$, where

$$f(\alpha) := \sum_{t=1}^T \sum_{j=0}^{T-t} \alpha^j = \sum_{j=0}^T (T-j)\alpha^j = \frac{\alpha^{T+1} - (T+1)\alpha + T}{(1-\alpha)^2}.$$

4.2. **Gaussian cash flows.** The convenient properties of conditional distributions of multivariate normal distributions allow for much stronger results than what have been possible in the setting considered so far. In what follows, we will therefore derive properties of the cost-of-capital margin in a Gaussian setting. Since the cost-of-capital margin is primarily intended for aggregate cash flows, Gaussian model assumptions will in many situations provide a reasonable approximation.

Definition 4. *Let Γ be a finite set of Gaussian vectors in $\mathbb{R}^T$ that are jointly Gaussian. Let*

$$\mathcal{G}_0 := \{\emptyset, \Omega\}, \quad \mathcal{G}_t := \Big(\vee_{Z \in \Gamma} \sigma(Z_t) \Big) \vee \mathcal{G}_{t-1} \quad \text{for } t = 1, \dots, T.$$

$\mathbb{G} := (\mathcal{G}_t)_{t=0}^T$ *is called a Gaussian filtration, and, if $X \in \Gamma$, then $(X, \mathbb{G})$ is called a Gaussian model.*

For a Gaussian model $(X, \mathbb{G})$, X is interpreted as a cash flow that may be assigned a value and $\mathbb{G}$ represents the flow of information used in the valuation of X. Notice that by Proposition 2 it is sufficient to only consider zero mean Gaussian cash flows X.

Consider a Gaussian model $(X, \mathbb{G})$ and let

$$(29) \qquad Y = a_0 + \sum_{Z \in \Gamma} \sum_{s=1}^{T} a_s^Z Z_s \text{ for some } a_0 \in \mathbb{R}, a_s^Z \in \mathbb{R}.$$

Then, the conditional distribution of Y given $\mathcal{G}_t$ is given by

$$(30) \qquad \mathbb{P}(Y \in A \mid \mathcal{G}_t) = \mathbb{P}(\mathbb{E}\left[Y \mid \mathcal{G}_t\right] + \text{Var}(Y \mid \mathcal{G}_t)^{1/2} \epsilon_{t+1} \in A \mid \mathcal{G}_t),$$

where ϵ_{t+1} is $\mathcal{G}_{t+1}$-measurable, standard normal and independent of $\mathcal{G}_t$. Moreover,

$$(31) \qquad \mathbb{E}\left[Y \mid \mathcal{G}_t\right] = b_0 + \sum_{Z \in \Gamma} \sum_{s=1}^{t} b_s^Z Z_s \text{ for some } b_0 \in \mathbb{R}, b_s^Z \in \mathbb{R},$$

and, a special feature of conditional Gaussian distributions that is essential here, $\text{Var}(Y \mid \mathcal{G}_s) \in L_+^0(\mathcal{G}_0)$. These properties ensure that if the mapping $W_t : L^p(\mathcal{G}_{t+1}) \to L^p(\mathcal{G}_t)$ given by (13), with $\mathbb{F} := \mathbb{G}$ and $\eta_t := \eta_0$, satisfying the assumptions in Proposition 4, and if $Y \in L^p(\mathcal{G}_{t+1})$ is of the form (29), then, by Proposition 4 and (31), for every $p \in [1, \infty)$,

$$W_t(Y) = \mathbb{E}\left[Y \mid \mathcal{G}_t\right] + \text{Var}(Y \mid \mathcal{G}_t)^{1/2} W_t(\epsilon_{t+1})$$
$$= \mathbb{E}\left[Y \mid \mathcal{G}_t\right] + \text{Var}(Y \mid \mathcal{G}_t)^{1/2} W_0(\epsilon_1).$$

Assumption 1. *For a zero mean Gaussian model $(X, \mathbb{G})$ and $p \in [1, \infty)$, $(W_t)_{t=0}^{T-1}$ is a sequence of mappings $W_t : L^p(\mathcal{G}_{t+1}) \to L^p(\mathcal{G}_t)$ as in Proposition 4 with $\mathbb{F} := \mathbb{G}$ and $\eta_t := \eta_0 \in L_+^0(\mathcal{G}_0)$. Moreover, ϵ_1 is $\mathcal{G}_1$-measurable and standard normal.*

Proposition 6. *Let $(X, \mathbb{G})$ be a zero mean Gaussian model and suppose that Assumption 1 holds. Then, for $t \in \{0, \dots, T-1\}$,*

$$V_{t, \mathbb{G}}(X) = \mathbb{E}\left[\sum_{s=t+1}^{T} X_s \mid \mathcal{G}_t \right] + \sum_{s=t+1}^{T} \text{Var}\left(\mathbb{E}\left[\sum_{u=s}^{T} X_u \mid \mathcal{G}_s \right] \mid \mathcal{G}_{s-1} \right)^{1/2} W_0(\epsilon_1).$$

Moreover,

$$V_{0,\mathbb{G}}(X) = \sum_{s=1}^{T} \Big(\operatorname{Var} \Big(\sum_{u=s}^{T} X_u \mid \mathcal{G}_{s-1} \Big) - \operatorname{Var} \Big(\sum_{u=s}^{T} X_u \mid \mathcal{G}_s \Big) \Big)^{1/2} W_0(\epsilon_1).$$

Notice that, given the assumptions of Proposition 6, we may express the cost-of-capital margin $V_{0,\mathbb{G}}(X)$ as

$$\sum_{s=1}^{T} \operatorname{Var} \Big(X_s + \mathbb{E} \Big[\sum_{u=s+1}^{T} X_u \mid \mathcal{G}_s \Big] - \mathbb{E} \Big[\sum_{u=s}^{T} X_u \mid \mathcal{G}_{s-1} \Big] \mid \mathcal{G}_{s-1} \Big)^{1/2} W_0(\epsilon_1).$$

In particular, the cost-of-capital margin $V_{0,\mathbb{G}}(X)$ is proportional to the sum of the conditional standard deviations of the errors of the repeated predictions of the sum of the remaining cash flows.

Proof of Proposition 6. The statement is proved by induction. Let $c := W_0(\epsilon_1)$ and $V_t(X) := V_{t,\mathbb{G}}(X)$. Clearly, X_T is of the form (29) so (30) holds. Therefore, from statement (iii) in Proposition 4,

$$V_{T-1}(X) := W_{T-1}(X_T) = \mathbb{E}\left[X_T \mid \mathcal{G}_{T-1} \right] + \operatorname{Var}(X_T \mid \mathcal{G}_{T-1})^{1/2} c.$$

Now let $t \leq T - 1$ and assume that the expression for $V_t(X)$ holds for t. Then $X_t + V_t(X)$ is of the form (29) so (30) holds. Therefore, from statement (iii) in Proposition 4,

$$\begin{aligned}
V_{t-1}(X) &:= W_{t-1}(X_t + V_t(X)) \\
&= \mathbb{E}\left[X_t + V_t(X) \mid \mathcal{G}_{t-1} \right] + \operatorname{Var}(X_t + V_t(X) \mid \mathcal{G}_{t-1})^{1/2} c \\
&= \mathbb{E}\left[\sum_{s=t}^{T} X_s \mid \mathcal{G}_{t-1} \right] + \sum_{s=t+1}^{T} \operatorname{Var}\Big(\mathbb{E}\Big[\sum_{u=s}^{T} X_u \mid \mathcal{G}_s \Big] \mid \mathcal{G}_{s-1} \Big)^{1/2} c \\
&\quad + \operatorname{Var}\Big(X_t + \mathbb{E}\Big[\sum_{s=t+1}^{T} X_s \mid \mathcal{G}_t \Big] \mid \mathcal{G}_{t-1} \Big)^{1/2} c \\
&= \mathbb{E}\left[\sum_{s=t}^{T} X_s \mid \mathcal{G}_t \right] + \sum_{s=t}^{T} \operatorname{Var}\Big(\mathbb{E}\Big[\sum_{u=s}^{T} X_u \mid \mathcal{G}_s \Big] \mid \mathcal{G}_{s-1} \Big)^{1/2} c.
\end{aligned}$$

Recall the variance decomposition formula: for $\mathcal{F} \subset \mathcal{G} \subset \mathcal{H}$ and $Y \in L^2(\mathcal{H})$,

$$\operatorname{Var}(Y \mid \mathcal{F}) = \mathbb{E}\left[\operatorname{Var}(Y \mid \mathcal{G}) \mid \mathcal{F} \right] + \operatorname{Var}(\mathbb{E}\left[Y \mid \mathcal{G} \right] \mid \mathcal{F}).$$

Applying the variance decomposition formula and using the fact that in the Gaussian case the conditional variance is a constant, we find that, for $s < T$,

$$\begin{aligned}
\operatorname{Var}\Big(&\mathbb{E}\Big[\sum_{u=s}^{T} X_u \mid \mathcal{G}_s \Big] \mid \mathcal{G}_{s-1} \Big) \\
&= \operatorname{Var}\Big(\sum_{u=s}^{T} X_u \mid \mathcal{G}_{s-1} \Big) - \mathbb{E}\Big[\operatorname{Var}\Big(\sum_{u=s}^{T} X_u \mid \mathcal{G}_s \Big) \mid \mathcal{G}_{s-1} \Big] \\
&= \operatorname{Var}\Big(\sum_{u=s}^{T} X_u \mid \mathcal{G}_{s-1} \Big) - \operatorname{Var}\Big(\sum_{u=s}^{T} X_u \mid \mathcal{G}_s \Big).
\end{aligned}$$

$\square$

Computation of $W_0(\epsilon_1)$ is illustrated in the following example.

Example 3. *Let ϵ_1 be standard normal with distribution and density function Φ and φ, respectively, and let $U_0(\cdot) = \mathbb{E}[\cdot]$. Then*

$$\begin{aligned}
W_0(\epsilon_1) &= R_0(-\epsilon_1) - \frac{1}{1+\eta_0}\mathbb{E}\left[(R_0(-\epsilon_1) - \epsilon_1)1_{\{\epsilon_1 \le R_0(-\epsilon_1)\}}\right] \\
&= R_0(-\epsilon_1) - \frac{1}{1+\eta_0}\Big(R_0(-\epsilon_1)\Phi(R_0(-\epsilon_1)) + \varphi(R_0(-\epsilon_1))\Big) \\
&\le R_0(-\epsilon_1)\frac{\eta_0}{1+\eta_0},
\end{aligned}$$

where the inequality is due to the Mill's ratio inequalities $\varphi(x)x/(1+x^2) \le 1 - \Phi(x) \le \varphi(x)/x$ for $x > 0$, see e.g. [10]. Notice that if $R_0 = \mathrm{VaR}_p$, then $R_0(-\epsilon_1) = \Phi^{-1}(1-p)$, and for Expected Shortfall, $R_0 = \mathrm{ES}_p$, corresponding to $m^R(u) = p^{-1}1_{[1-p,1]}(u)$ in Example 2, $R_0(-\epsilon_1) = p^{-1}\varphi(\Phi^{-1}(1-p))$.

Bounds on the cost-of-capital margin can be obtained from Proposition 6. From the proof of Proposition 6, notice that

$$(32) \qquad V_{t,\mathbb{G}}(X) - \mathbb{E}\left[\sum_{s=t+1}^{T} X_s \mid \mathcal{G}_t\right] = W_0(\epsilon_1)\sum_{s=t+1}^{T} a_s^{1/2},$$

where, for $s = 1, \ldots, T$,

$$(33) \qquad a_s := \mathrm{Var}\Big(\sum_{u=t+1}^{T} X_u \mid \mathcal{G}_{s-1}\Big) - \mathrm{Var}\Big(\sum_{u=t+1}^{T} X_u \mid \mathcal{G}_s\Big).$$

In particular,

$$\sum_{s=t+1}^{T} a_s = \mathrm{Var}\Big(\sum_{s=t+1}^{T} X_s \mid \mathcal{G}_t\Big).$$

An upper bound on the left-hand side in (32) is found by solving a standard convex optimization problem:

$$\begin{aligned}
&\text{maximize} \quad \sum_{s=t+1}^{T} a_s^{1/2} \\
&\text{subject to} \quad \sum_{s=t+1}^{T} a_s = C, \quad a_s \ge 0
\end{aligned}$$

The concave objective function has a unique maximum for $a_s = C/(T-t)$ for all s. Minimizing the objective function over the same (convex) set gives a minimum for $a_{s_0} = C$ for some $s_0 \in \{t+1, \ldots, T\}$ and $a_s = 0$ for $s \ne s_0$. We have thus proved the following bounds.

Proposition 7. *Let $(X, \mathbb{G})$ be a zero mean Gaussian model and suppose that Assumption 1 holds. Then,*

$$W_0(\epsilon_1) \operatorname{Var} \Big(\sum_{s=t+1}^{T} X_s \mid \mathcal{G}_t \Big)^{1/2} \leq V_{t,\mathbb{G}}(X) - \mathbb{E} \left[\sum_{s=t+1}^{T} X_s \mid \mathcal{G}_t \right]$$

$$\leq W_0(\epsilon_1)(T-t)^{1/2} \operatorname{Var} \Big(\sum_{s=t+1}^{T} X_s \mid \mathcal{G}_t \Big)^{1/2}.$$

In particular,

$$W_0(\epsilon_1) \operatorname{Var} \Big(\sum_{s=1}^{T} X_s \Big)^{1/2} \leq V_{0,\mathbb{G}}(X) \leq W_0(\epsilon_1) T^{1/2} \operatorname{Var} \Big(\sum_{s=1}^{T} X_s \Big)^{1/2}.$$

The interpretation of the upper bound for the cost-of-capital margin $V_{0,\mathbb{G}}(X)$ in Proposition 7 is as follows. The Gaussian model $(X, \mathbb{G})$ maximizing $V_{0,\mathbb{G}}(X)$ corresponds to a filtration $\mathbb{G}$ such that

$$\operatorname{Var} \Big(\sum_{u=1}^{T} X_u \mid \mathcal{G}_{s-1} \Big) - \operatorname{Var} \Big(\sum_{u=1}^{T} X_u \mid \mathcal{G}_s \Big) = \frac{1}{T} \operatorname{Var} \Big(\sum_{u=1}^{T} X_u \Big),$$

i.e. the uncertainty (variance) in the remaining cash flow is distributed evenly over the length of the cash flow.

The interpretation of the lower bound in Proposition 7 is as follows. Given a zero mean Gaussian cash flow X, such that $\operatorname{Var}(X_1) > 0$, the Gaussian model $(X, \mathbb{G})$ minimizing $V_{0,\mathbb{G}}(X)$ corresponds to a filtration $\mathbb{G}$ such that

$$\operatorname{Var} \Big(\sum_{u=1}^{T} X_u \mid \mathcal{G}_s \Big) = 0 \quad \text{for } s \geq 1$$

i.e. the cash flow after time 1 is completely known at time 1. In particular, after time 1 there is no need for capital funds and hence there are no capital costs. This interpretation follows from Proposition 8 below.

Proposition 8. *Let $(X, \mathbb{G})$ and $(X, \widetilde{\mathbb{G}})$ be two Gaussian models such that, for every t, $\widetilde{\mathcal{G}}_t = \mathcal{G}_s$ for some $s \geq t$. Suppose further that, for both Gaussian models, Assumption 1 holds. Then $V_{0,\mathbb{G}}(X) \geq V_{0,\widetilde{\mathbb{G}}}(X)$.*

Proof. It is sufficient to consider the case $\widetilde{\mathcal{G}}_{t_0} = \mathcal{G}_{t_0+1}$ for some $t_0 \geq 1$, and $\widetilde{\mathcal{G}}_t = \mathcal{G}_t$ for $t \neq t_0$. Repeating the argument then yields the conclusion. Set $b_t := \operatorname{Var}(\sum_{u=1}^{T} X_u \mid \mathcal{G}_t)$, $\widetilde{b}_t := \operatorname{Var}(\sum_{u=1}^{T} X_u \mid \widetilde{\mathcal{G}}_t)$, $a_t := b_{t-1} - b_t$ and $\widetilde{a}_t := \widetilde{b}_{t-1} - \widetilde{b}_t$. Then

$$
\begin{aligned}
V_{0,\mathbb{G}}(X) - V_{0,\widetilde{\mathbb{G}}}(X) &= W_0(\epsilon_1) \sum_{t=1}^{T} (a_t^{1/2} - \widetilde{a}_t^{1/2}) \\
&= W_0(\epsilon_1) \Big(a_{t_0}^{1/2} + a_{t_0+1}^{1/2} - \widetilde{a}_{t_0}^{1/2} - \widetilde{a}_{t_0+1}^{1/2} \Big) \\
&= W_0(\epsilon_1) \Big(a_{t_0}^{1/2} + a_{t_0+1}^{1/2} - 0 - (a_{t_0} + a_{t_0+1})^{1/2} \Big) \\
&\geq 0
\end{aligned}
$$

due to the subadditivity of $\mathbb{R}_+ \ni x \mapsto x^{1/2} \in \mathbb{R}_+$. $\qquad \square$

In the Gaussian setting, the cost-of-capital margin is subadditive. If the aggregate liability cash flow is decomposed into a sum of sub-liability cash flows, then the sum of the corresponding cost-of-capital margins dominates the cost-of-capital margin for the aggregate liability cash flow.

Proposition 9. *Let $((X, \widetilde{X}), \mathbb{G})$ be a Gaussian model and suppose that Assumption 1 holds. Then, for $t \in \{0, \ldots, T-1\}$, $V_{t,\mathbb{G}}(X + \widetilde{X}) \leq V_{t,\mathbb{G}}(X) + V_{t,\mathbb{G}}(\widetilde{X})$.*

Proof. From Proposition 6,

$$W_0(\epsilon_1)^{-1}\Big(V_{t,\mathbb{G}}(X + \widetilde{X}) - V_{t,\mathbb{G}}(X) - V_{t,\mathbb{G}}(\widetilde{X})\Big) = \sum_{s=t+1}^{T} \Delta_{s,T},$$

where

$$\Delta_{s,T} = \mathrm{Var}\left(\mathbb{E}\left[\sum_{u=s}^{T} X_u \mid \mathcal{G}_s\right] + \mathbb{E}\left[\sum_{u=s}^{T} \widetilde{X}_u \mid \mathcal{G}_s\right] \mid \mathcal{G}_{s-1}\right)^{1/2}$$
$$- \mathrm{Var}\left(\mathbb{E}\left[\sum_{u=s}^{T} X_u \mid \mathcal{G}_s\right] \mid \mathcal{G}_{s-1}\right)^{1/2} - \mathrm{Var}\left(\mathbb{E}\left[\sum_{u=s}^{T} \widetilde{X}_u \mid \mathcal{G}_s\right] \mid \mathcal{G}_{s-1}\right)^{1/2}$$

The conclusion now follows from the general fact

$$\mathrm{Var}(Z + \widetilde{Z} \mid \mathcal{G}) \leq (\mathrm{Var}(Z \mid \mathcal{G})^{1/2} + \mathrm{Var}(\widetilde{Z} \mid \mathcal{G})^{1/2})^2.$$

$\square$

5. Valuation of a life-insurance portfolio

We will now go through a simple, yet realistic, life-insurance example aimed at illustrating aspects of the cost-of-capital margin.

One of the simplest insurance contracts that is non-trivial and for which it is possible to carry out an exact valuation is a portfolio which at time 0 consists of n_0 identical and independent term-life-insurance contracts. A term-life-insurance contract for a today x years old individual that terminates at latest T years from today is constructed as follows: if the insured individual

- dies during year $t+1$ (between times t and $t+1$) for $t \in \{0, \ldots, T-1\}$, the amount 1 is paid to the beneficiary at time $t + 1$ (payment at end of year),
- is alive after T years, the contract pays nothing.

Given the above, at time 0 there are $N_0 = n_0$ active contracts of x years old individuals terminating at time T. Moreover, deaths of individuals are assumed to be independent events. Further, if we let D_{t+1} denote the number of deaths during year $t+1$, the dynamics of the number of active contracts at each time can be described as a nested binomial process as follows: $(N_t)_{t=0}^{T}$ is a Markov process with $N_{t+1} := N_t - D_{t+1}$ and $D_{t+1} | N_t \sim \mathrm{Bin}(N_t, q_{x+t})$. Here

$$q_{x+t} := \mathbb{P}(T_x \leq t + 1 | T_x > t) = 1 - \frac{S_x(t+1)}{S_x(t)}, \quad t = 0, \ldots, T-1,$$

where T_x is the remaining lifetime of a today x years old individual, where

$$S_x(u) := \exp\left\{ -\int_x^{x+u} \mu_s ds \right\},$$

where $\mu_s \geq 0$ is the so-called mortality law or force of mortality. Here we let μ_s be the Makeham mortality law given by

$$\mu_x := \alpha + \beta \exp\{\gamma x\}, \ \alpha, \beta, \gamma > 0.$$

In the numerical calculations carried out below we will use $\alpha = 0.001, \beta = 0.000012$ and $\gamma = 0.101314$, corresponding to Swedish mortality table M90 for males. The q_{x+t}s are so-called deferred death probabilities, sometimes denoted by $_{t|1}q_x$. Appendix A describes how the recursive valuation is formulated for a homogeneous population.

If we instead would assign a value to this liability cash flow using EIOPA's standard procedure, we would let the liability value correspond to the so-called Technical Provisions (TP) given by

$$\mathrm{TP}(D) := \mathrm{BE}_{\mu,1}(D) + \mathrm{RM}(D),$$

where,

$$\mathrm{BE}_{\mu,i}(D) := \sum_{j=i}^{T} \frac{\mathbb{E}\left[D_j; \mu\right]}{(1 + r_{i-1,j})^{j-i+1}}, \quad i = 1, \dots, T,$$

where $\mathbb{E}\left[D_j; \mu\right]$ denotes expected value of D_j using mortality law μ and $r_{i,j}$ is the yearly compounded forward rate for the period i to j. That is, if we let s_i denote the yearly compounded spot rate from 0 to i, it holds that $(1+s_{i+1})^{i+1} = (1+s_i)^i(1+r_{i,i+1})$. Typically the s_is denotes EIOPA's annual zero-coupon spot rates where $s_i := r_{0,i}$. Further,

$$\mathrm{SCR}(D) := \mathrm{BE}_{\widetilde{\mu},1}(D) - \mathrm{BE}_{\mu,1}(D),$$

$\widetilde{\mu} = 1.15\mu$ corresponds to the stressed mortality defined by EIOPA, see [9, Article 137], and the risk margin RM is finally given by

$$\mathrm{RM}(D) := \mathrm{CoC} \, \frac{\mathrm{SCR}(D)}{\mathrm{BE}_{\mu,1}(D)} \sum_{i=1}^{T} \frac{\mathrm{BE}_{\mu,i}(D)}{(1 + r_{0,i})^i},$$

see [8, Paragraph 1.114, Method 2]. In the examples below EIOPA's annual zero-coupon spot rates are the basis of all discounting, and in particular we use the Swedish curve as of September 30th 2016. Further, CoC is the cost-of-capital rate, taken to be 0.06. Computations of $\mathrm{BE}_{\widetilde{\mu},1}(D)$ and $\mathrm{BE}_{\mu,i}(D)$ are found in Appendix A.

Note that the above definition of the BE together with how the discounting is carried out for future BE will result in a net effect of discounting in the (double)sum defining RM corresponding to discounting once with respect to the original spot curve provided by EIOPA. This is the only way we can make any sense of the definition of SCR and RM, i.e. avoid "double" discounting.

Figure 1 shows BE as a function of time to contract expiry for a portfolio consisting of $n_0 = 1000$ 50 years old Swedish males.

Figure 1 shows a comparison between the valuation according to our interpretation of EIOPA's standard valuation procedure for a portfolio consisting
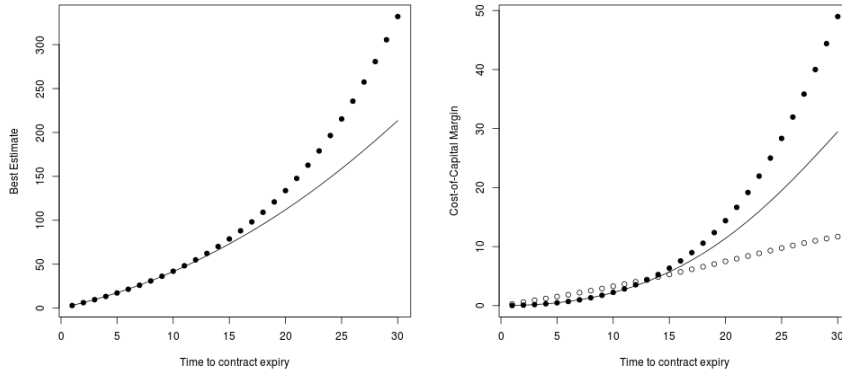
FIGURE 1. The left figure shows best estimate as function of time to contract expiry for a term-life-insurance portfolio consisting of $n_0 = 1000$ independent 50 years old Swedish males. For the same portfolio, the right figure compares the cost-of-capital margins for the nested binomial model (circles) and EIOPA risk margins *without discounting* (solid black discs). The solid black lines corresponds to the standard formula calculations *with* discounting.

of $N_0 = n_0 = 1000$ identical 50 years old Swedish males, and the valuation according to (13) and (18). Here R_t is set to conditional VaR at the 0.5% level, seen as a mapping $R_t : L^1(\mathcal{F}_{t+1}) \to L^1(\mathcal{F}_t)$ with $\mathbb{F}$ taken to be the filtration generated by (N_t), $U_t := \mathbb{E}[\cdot \mid \mathcal{F}_t]$, and $\eta_t := \mathrm{CoC} := 0.06$. Information about computational aspects can be found in Appendix A.1. Each circle in the left panel of Figure 1 corresponds to a specific (total) best estimate for a portfolio consisting of 1000 contracts of 50 years old Swedish males that had the same specific time to contract expiry at the time of valuation (time 0). That is, each point corresponds to a new portfolio with a certain time to expiry and not the evolution of *one* portfolio with duration 30 years. This interpretation applies to all figures below.

For the sake of comparison we focus on EIOPA's risk margin together with the value of the residual cash flow as defined in Section 3. From Figure 1 it can be seen that in this situation EIOPA's risk margin may underestimate as well as overestimate the risk compared to the above more correct valuation procedure. One can argue that the EIOPA method used here is an approximation, but it does not necessarily seem to be a prudent one. This is unfortunate, since the authors believe that this method is commonly used in the industry, given that the so-called proportionality principle applies. Further, from Figure 1 it is seen that, when discounting is used, EIOPA's risk margin is closer to the cost-of-capital margin. Recall that the cost-of-capital margin is the second step in the two-step valuation procedure, where discounting appears in the first step (value of replicating portfolio) but not in the cost-of-capital margin valuation applied to the residual liability cash flow. Consequently, discounting will affect EIOPA's method, but not the cost-of-capital margin. Further, even in the low interest rate environment

as of September 30th 2016, the effect of discounting is still substantial for moderate portfolio durations. Moreover, the observed differences between EIOPA's method and the method proposed in the present paper are in general hard to understand, due to that EIOPA's method is ad hoc.

Further, one can note that the recursions defined in Appendix A.1 are expressed for a single homogeneous population. For heterogeneous populations, numerical problems with computation of binomial probabilities may arise. Therefore, it is of interest to analyse how well the Gaussian approximation of Section 4.2 performs, since this can be readily adapted to handle heterogeneous populations. In order to do so, we want to compare the cash flow generated by D with that of $\mathbb{E}[D] + X$, where X is a zero mean Gaussian vector with the same covariance matrix as D, following the setup of Section 4.2. Detailed calculations showing how this is done can be found in Appendix A.2 and A.3.
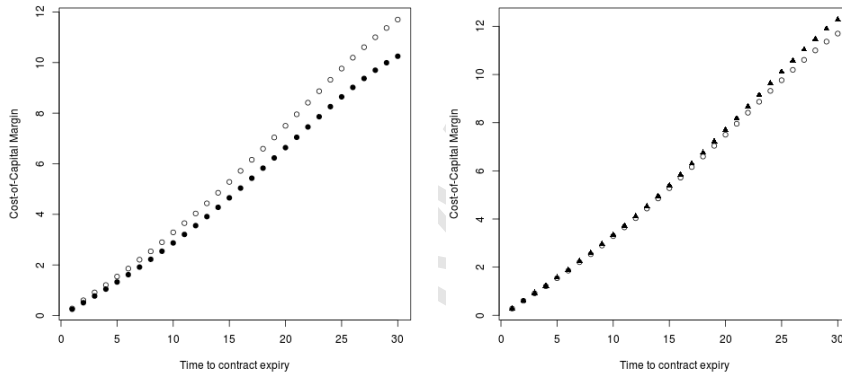

FIGURE 2. For a term-life-insurance portfolio consisting of $n_0 = 1000$ independent 50 years old Swedish males, the left figure compares cost-of-capital margins for the nested binomial model (circles) and the Gaussian approximation (solid black discs). The right figure compares cost-of-capital margins for the nested binomial model (circles) and the upper bound (21) (solid black triangles).

In Figure 2 we see that the approximation is performing well for the chosen times to expiry of the insurance contracts, given a portfolio size of $n_0 = 1000$. For larger portfolio sizes and more complex insurance products the Gaussian approximation will serve as a natural benchmark method. Further, from the panel on the right-hand side of Figure 2 it is also seen that the upper bound given by (21) is close to the exact calculation.

Even though the Gaussian approximation is close to the correct valuation for this particular portfolio of term-life contracts, it is seen that the approximation yields an underestimation. Another simple type of insurance contract is given by the analogous temporary annuity contract that pays an amount 1 at the end of each year when the insured is alive, but at most for T years. In analogy with the term-life contracts above, the best estimate

from year i for the temporary annuities is given by

$$\mathrm{BE}_{\mu,i}(N) := \sum_{j=i+1}^{T} \frac{\mathbb{E}\left[N_j;\mu\right]}{(1+r_{i-1,j})^{j-i+1}}, \quad i = 1, \ldots, T,$$

where $\mathbb{E}\left[N_j;\mu\right]$ denotes expected value of N_j using mortality law μ, and the expressions for SCR etc. follow verbatim with the exception that the stress is a 15% decrease. The calculations for the corresponding number of contracts active at the end of each calendar year follow the ones carried out for the term-life contracts, see Appendix A.
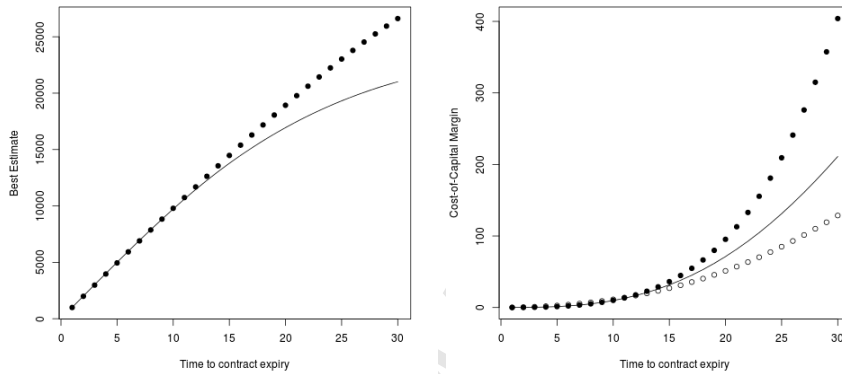

FIGURE 3. The left figure shows best estimate as function of time to contract expiry for a temporary annuity portfolio consisting of $n_0 = 1000$ independent 50 years old Swedish males. For the same portfolio, the right figure compares the cost-of-capital margins for the nested binomial model (circles) and EIOPA risk margins *without discounting* (solid black discs). The solid black lines corresponds to the standard formula calculations *with* discounting.

Figure 3 shows the the analog of Figure 1 for a portfolio of $n_0 = 1000$ independent temporary annuities for 50 years old Swedish males. Again it is seen that EIOPA's method for most times to expiry produces larger values than the cost-of-capital margin, but not for the shorter times to expiry.

Regarding the Gaussian approximation, it is seen from Figure 4 that in the situation with a portfolio of temporary annuities the Gaussian approximation produces prudent estimates that are very close to the upper bound given by (21).
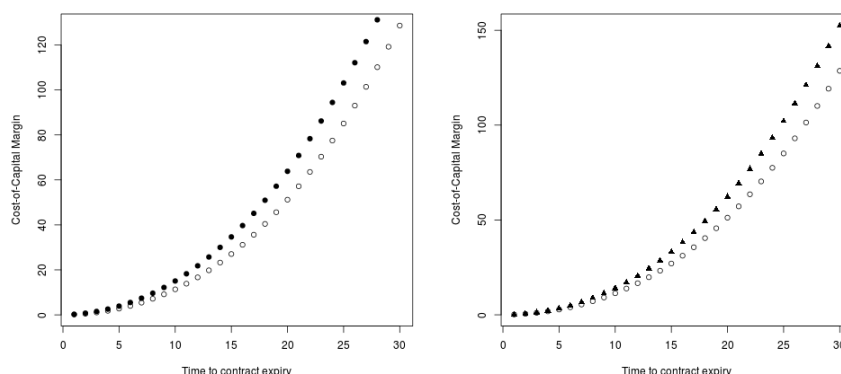
FIGURE 4. For a temporary annuity portfolio consisting of $n_0 = 1000$ independent 50 years old Swedish males, the left figure compares cost-of-capital margins for the nested binomial model (circles) and the Gaussian approximation (solid black discs). The right figure compares cost-of-capital margins for the nested binomial model (circles) and the upper bound (21) (solid black triangles).

REFERENCES

[1] Carlo Acerbi (2002), Spectral measures of risk: A coherent representation of subjective risk aversion. *Journal of Banking & Finance*, 26, 1505-1518.

[2] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber and David Heath (1999), Coherent measures of risk. *Mathematical Finance* 4, 203-228.

[3] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, David Heath and Hyejin Ku (2007), Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research*, 152, 5-22.

[4] Patrick Cheridito, Michael Kupper and Freddy Delbaen (2006), Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability*, 11, 57-106.

[5] Patrick Cheridito and Mitja Stadje (2009), Time-inconsistency of VaR and time-consistent alternatives. *Finance Research Letters*, 6(1), 40-46.

[6] Jaksa Cvitanic and Ioannis Karatzas (1999), On dynamic measures of risk. *Finance & Stochastics*, 3(4), 451-482.

[7] Jan Dhaene, Alexander Kukush, Daniël Linders, Qihe Tang (2012), Remarks on quantiles and distortion risk measures. *European Actuarial Journal* 2, 319-328.

[8] EIOPA (2015), Guidelines on the valuation of technical provisions - EIOPA-BoS-14/166 EN. *EIOPAs official website*.

[9] European Commission (2015), Commission Delegated Regulation (EU) 2015/35 of 10 October 2014. *Official Journal of the European Union*.

[10] R. D. Gordon (1941), Values of Mill's ratio of area to bounding ordinate of the normal probability integral for large values of the argument. *Annals of Mathematical Statistics*, 12, 364-366.

[11] Christoph Möhr (2011), Market-consistent valuation of insurance liabilities by cost of capital. *ASTIN Bulletin*, 41, 315-341.

[12] Alfred Müller (2007), Certainty equivalents as risk measures. *Brazilian Journal of Probability and Statistics*, 21, 1-12.

[13] Esbjörn Ohlsson and Jan Lauzeningks (2008), The one-year non-life insurance risk. Conference Paper, *ASTIN Colloquium 2008*, Manchester.

[14] Georg Ch. Pflug and Andrzej Ruszczyński (2003), A risk measure for income streams. in Risk Measures for the 21st Century, G. Szegoe (Ed.). J. Wiley and Sons, 2003.
[15] Georg Ch. Pflug and Andrzej Ruszczyński (2005), Measuring risk for income streams. Computational Optimization and Applications, 32(1), 161-178.
[16] Frank Riedel (2004), Dynamic coherent risk measures. *Stochastic Processes and their Applications*, 112, 185-200.
[17] Robert Salzmann and Mario V. Wüthrich (2010), Cost-of-capital margin for a general insurance liability runoff. *ASTIN Bulletin* 40(2), 415-451.
[18] Leonard J. Savage, *The Foundations of Statistics*. Wiley, 1954.
[19] Alexander Shapiro (2012), Time consistency of dynamic risk measures. *Operations Research Letters* 40(6), 436-439.
[20] Andreas Tsanakas, Mario V. Wüthrich and Aleš Černý (2013), Market value margin via mean-variance hedging. *ASTIN Bulletin* 43(3), 301-322.
[21] Shaun S. Wang (1996), Premium calculation by transforming the layer premium density. *ASTIN Bulletin* 26, 71-92.
[22] Shaun S. Wang and Virginia R. Young (1998), Ordering risks: Expected utility theory versus Yaari's dual theory of risk. *Insurance: Mathematics and Economics* 22, 145-161.
[23] Tan Wang (1999), A class of dynamic risk measures. Working paper, U.B.C.
[24] Mario V. Wüthrich, Paul Embrechts and Andreas Tsanakas (2011), Risk margin for a non-life insurance run-off. *Statistics & Risk Modeling* 28, 299-317.
[25] Mario V. Wüthrich and Michael Merz, *Financial Modeling, Actuarial Valuation and Solvency in Insurance*, Springer, 2013.
[26] Menahem E. Yaari (1987), The dual theory of choice under risk. *Econometrica* 55(1), 95-115.

APPENDIX A. COMPUTATIONAL DETAILS FOR SECTION 5

A.1. **Recursive valuation for a homogeneous population.** Let W_t be defined as in Section 3. The backward recursive valuation for the nested binomial model of Section 5 can be expressed explicitly as follows. Write $V_t(D) := G_t(N_t)$, where N_t denotes the number of active contracts at time t, and G_t is some deterministic function, presuming that $\mathcal{F}_t = \sigma(N_1, ..., N_t)$. With this notation,

$$Y_{t+1} = D_{t+1} + G_{t+1}(N_{t+1}) = D_{t+1} + G_{t+1}(N_t - D_{t+1})$$

and we get the following recursion formula for all $n \in \{0, 1, ..., N_0\}$: $G_T(n) = 0$ and, for $t = 0, \ldots, T - 1$,

$$G_t(n) = W_t(D_{t+1} + G_{t+1}(n - D_{t+1})), \quad D_{t+1} \sim \text{Bin}(n, q_t),$$

where the dependence on the age of the insured population has been omitted to simplify the notation. Starting by calculating $G_{T-1}(n)$ for all feasible values of n, we can use the recursive formula until we reach $V_0(D) = G_0(N_0)$. The computational feasibility relies on the fact that the relevant information at time t is contained in N_t which takes values in the relatively small set $\{0, \ldots, N_0\}$. If we consider information which may be expressed as a vector in $\mathbb{N}^k$, $k \geq 2$, direct computation will be considerably more involved, and possibly unfeasible. This would be the case if the population of insured consisted of $k \geq 2$ homogeneous subgroups.

The above expressions describe the valuation recursions used to value homogeneous term-life portfolios, but the same reasoning applies to obtain valuation recursions for temporary annuity portfolios.

A.2. **Computation of moments in the nested binomial model.** We will now go through how $\mathbb{E}[D_i]$ and $\mathbb{E}[D_i D_j]$ are calculated for the nested binomial model described in Section 5. In order to ease notation we omit explicit references to the age x of the insured population. First, note that with $q_k := \mathbb{P}(T_x \leq k+1 \mid T_x > k)$,

$$N_i \sim \mathrm{Bin}(n_0, \widetilde{p}_i), \quad \widetilde{p}_i := \prod_{k=0}^{i-1}(1-q_k), \quad i \geq 1,$$

$$D_{i+1} \mid N_l \sim \mathrm{Bin}(N_l, q_{i|l}), \quad q_{i|l} := q_i \prod_{k=l}^{i-1}(1-q_k), \quad i \geq l, l \geq 0.$$

In particular, $\mathbb{E}[D_{i+1}] = q_{i|0} n_0$, which by repeated conditioning yields

$$\mathbb{E}[D_{i+1}] = n_0 q_i \tilde{p}_i = n_0 \mathbb{P}(i < T_x \leq i+1),$$
$$\mathrm{Var}(D_{i+1}) = n_0 q_i \tilde{p}_i (1 - q_i \tilde{p}_i).$$

Moreover, for $j > i \geq 0$, we have that

$$\mathbb{P}(D_{i+1} = y, D_{j+1} = z, N_i = n)$$
$$= \mathbb{P}(D_{i+1} = y, D_{j+1} = z \mid N_i = n)\mathbb{P}(N_i = n)$$

where

$$\mathbb{P}(D_{i+1} = y, D_{j+1} = z \mid N_i = n)$$
$$= \mathbb{P}(D_{i+1} = y \mid N_i = n)\mathbb{P}(D_{j+1} = z \mid D_{i+1} = y, N_i = n)$$
$$= \mathbb{P}(D_{i+1} = y \mid N_i = n)\mathbb{P}(D_{j+1} = z \mid N_{i+1} = n - y)$$
$$= \binom{n}{y} q_i^y (1 - q_i)^{n-y} \binom{n-y}{z} q_{j|(i+1)}^z (1 - q_{j|(i+1)})^{n-y-z}.$$

Combining these expressions, we get

$$\mathbb{P}(D_{i+1} = y, D_{j+1} = z, N_i = n)$$
$$= \binom{n}{y} q_i^y (1 - q_i)^{n-y} \binom{n-y}{z} q_{j|(i+1)}^z (1 - q_{j|(i+1)})^{n-y-z} \binom{n_0}{n} \widetilde{p}_i^n (1 - \widetilde{p}_i)^{n_0-n}$$
$$= \binom{n_0}{n_0-n, y, z, n-y-z}(1 - \tilde{p}_i)^{n_0-n}(q_i \tilde{p}_i)^y (q_j \tilde{p}_j)^z (\tilde{p}_{i+1}(1 - q_{j|(i+1)}))^{n-y-z},$$

which corresponds to

$$(D_{i+1}, D_{j+1}, N_i) =_d (\tilde{D}_i, D_{i+1}, D_{j+1}, M_{i,j}),$$

where $(\tilde{D}_i, D_{i+1}, D_{j+1}, M_{i,j}) \sim \mathrm{Mult}(n_0; 1 - \tilde{p}_i, q_i \tilde{p}_i, q_j \tilde{p}_j, \tilde{p}_{i+1}(1 - q_{j|(i+1)}))$ with $\tilde{D}_i$ and $M_{i,j}$ being auxiliary random variables. The multinomial interpretation then immediately gives

$$\mathrm{Cov}(D_{i+1}, D_{j+1}) = n_0 q_i \tilde{p}_i q_j \tilde{p}_j$$
$$= n_0 \mathbb{P}(i < T_x \leq i+1)\mathbb{P}(j < T_x \leq j+1),$$

and it may be verified that $\mathbb{E}[D_{i+1}]$ and $\mathrm{Var}(D_{i+1})$ calculated using this approach agrees with the expressions from above.

Similar calculations as those above can be carried out for N_i which yields the relevant expressions for temporary annuity portfolios.

A.3. **Cost-of-capital recursion for Gaussian models.** Here we derive an explicit recursion formula for the cost-of-capital margin for the Gaussian model $(X, \mathbb{G})$ with $\mathbb{G} = (\mathcal{G}_t)_{t=0}^T$, $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and $\mathcal{G}_t = \sigma(X_1, ..., X_t)$ for $t = 1, \ldots, T$. A similar, albeit more complicated, formula could be derived for a general Gaussian model, meaning that $\mathbb{G}$ is larger than the natural filtration of X.

For a multivariate normal vector $Z \sim N_n(\mu, \Sigma)$, write

$$\mu = \left[\begin{array}{c} \mu_{1:n-1} \\ \mu_n \end{array} \right], \quad \Sigma = \left[\begin{array}{cc} \Sigma_{1:n-1,1:n-1} & \Sigma_{1:n-1,n} \\ \Sigma_{n,1:n-1} & \Sigma_{n,n} \end{array} \right]$$

It is well known that the conditional distribution of Z_n given $Z_1, \ldots, Z_{n-1}$ is normal with parameters

$$\mu_{n|1:n-1} = \mu_n + \Sigma_{n,1:n-1} \Sigma_{1:n-1,1:n-1}^{-1} (Z_{1:n-1} - \mu_{1:n-1}),$$
$$\Sigma_{n|1:n-1} = \Sigma_{n,n} - \Sigma_{n,1:n-1} \Sigma_{1:n-1,1:n-1}^{-1} \Sigma_{1:n-1,n}.$$

Proposition 10. *Let $X = (X_t)_{t=1}^T \sim N_T(0, \Sigma)$ where Σ is invertible, let $\mathbb{G}$ be its natural filtration, and suppose that Assumption 1 holds. Then*

$$V_t(X) = \left\{ \begin{array}{ll} 0, & t = T, \\ (v^{(t)})^{\mathrm{T}} X_{1:t} + k_t, & t \in \{1, \ldots, T-1\}, \\ k_0, & t = 0, \end{array} \right.$$

where $k_T := 0$, $v^{(T)} := 0$, $v^{(0)} := 0$, and, for $t \in \{1, \ldots, T-1\}$, $k_t \in \mathbb{R}$ and $v_t \in \mathbb{R}^t$ can be calculated recursively from

$$k_t := k_{t+1} + W_t((1 + v_{t+1}^{(t+1)}) \Sigma_{t+1|1:t}^{1/2} \epsilon_{t+1}),$$
$$(v^{(t)})^{\mathrm{T}} := (v_{1:t}^{(t+1)})^{\mathrm{T}} + (1 + v_{t+1}^{(t+1)}) \Sigma_{t+1:1:t} \Sigma_{1:t,1:t}^{-1},$$

where $(\epsilon_t)_{t=1}^T$ is a sequence of independent standard normally distributed random variables such that ϵ_{t+1} is $\mathcal{G}_{t+1}$-measurable and independent of $\mathcal{G}_t$.

Proof of Proposition 10. We know that $V_T(X) = 0$ and we set $v^{(T)} := 0$. We prove the statement via induction. Take $t \in \{0, \ldots, T-1\}$ and suppose that $V_{t+1}(X) = (v^{(t+1)})^{\mathrm{T}} X_{1:t+1} + k_{t+1}$. First, consider the case $t \geq 1$. Then

$$Y_{t+1} = (1 + v_{t+1}^{(t+1)}) X_{t+1} + (v_{1:t}^{(t+1)})^{\mathrm{T}} X_{1:t} + k_{t+1}.$$

Since the latter two terms are $\mathcal{G}_t$-measurable, translation invariance of W_t combined with properties of the conditional Gaussian distribution give

$$\begin{aligned} V_t(X) &= W_t(Y_{t+1}) \\ &= W_t((1 + v_{t+1}^{(t+1)}) X_{t+1}) + (v_{1:t}^{(t+1)})^{\mathrm{T}} X_{1:t} + k_{t+1} \\ &= W_t((1 + v_{t+1}^{(t+1)})(\mu_{t+1|1:t} + \Sigma_{t+1|1:t}^{1/2} \epsilon_{t+1})) + (v_{1:t}^{(t+1)})^{\mathrm{T}} X_{1:t} + k_{t+1} \\ &= W_t((1 + v_{t+1}^{(t+1)}) \Sigma_{t+1|1:t}^{1/2} \epsilon_{t+1}) + (1 + v_{t+1}^{(t+1)}) \mu_{t+1|1:t} + (v_{1:t}^{(t+1)})^{\mathrm{T}} X_{1:t} + k_{t+1} \\ &= \left((v_{1:t}^{(t+1)})^{\mathrm{T}} + (1 + v_{t+1}^{(t+1)}) \Sigma_{t+1,1:t} \Sigma_{1:t,1:t}^{-1} \right)^{\mathrm{T}} X_{1:t} \\ &\quad + k_{t+1} + W_t((1 + v_{t+1}^{(t+1)}) \Sigma_{t+1|1:t}^{1/2} \epsilon_{t+1}) \\ &= (v^{(t)})^{\mathrm{T}} X_{1:t} + k_t \end{aligned}$$

from which the conclusion follows. Finally, consider the case $t = 0$. Then

$$
\begin{aligned}
V_0(X) &= W_0(Y_1) \\
&= W_0((1 + v_1^{(1)})X_1) + k_1 \\
&= k_1 + (1 + v_1^{(1)})\mu_1 + W_0((1 + v_1^{(1)})\Sigma_{1,1}^{1/2}\epsilon_1) \\
&= k_0.
\end{aligned}
$$

$\square$

Notice that $W_t(K\epsilon_{t+1}) = |K|W_0(\epsilon_1)$ due to symmetry of the standard normal distribution and Assumption 1.

(Hampus Engsner) Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden

E-mail address: hampus.engsner@math.su.se

(Mathias Lindholm) Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden

E-mail address: lindholm@math.su.se

(Filip Lindskog) Department of Mathematics, Stockholm University, SE-106 91 Stockholm, Sweden

E-mail address, corresponding author: lindskog@math.su.se