Astin Bulletin (2026), **56**, pp. 420–446
doi:10.1017/asb.2026.10084

International Actuarial Association
Association Actuarielle Internationale

RESEARCH ARTICLE

Hedging targeted risks with reinforcement learning: application to life insurance contracts with embedded guarantees

Carlos Octavio Pérez-Mendoza and Frédéric Godin

Concordia University, Canada
Corresponding author: Frédéric Godin; Email: frederic.godin@concordia.ca

Received: 2 June 2025; **Revised:** 31 October 2025; **Accepted:** 22 December 2025; **First published online:** 20 February 2026

Keywords: Deep hedging; reinforcement learning; risk allocation; Shapley decomposition; variable annuities

JEL codes: C45; C61; G22; G32

Abstract

We propose a deep reinforcement learning (RL) framework designed to optimize the hedging of specific, user-defined risk factors—referred to as targeted risks—in financial instruments affected by multiple sources of uncertainty. Our methodology uses Shapley value decompositions to establish source of risk grouping's contribution to the projected contract cash flows, providing a clear attribution of the profit and loss to distinct risk categories. Leveraging this decomposition, we apply deep RL to hedge only the targeted risks, while leaving non-targeted risks mostly unaffected. In addition, we introduce a joint neural network architecture in which the agent network utilizes risk estimates from a risk measurement neural network to stabilize the hedging strategy, taking into account local risk dynamics. Numerical experiments show that our approach outperforms traditional methods, such as delta hedging and traditional deep hedging, significantly reducing targeted risks in variable annuities while maintaining flexibility for broader applications.

1. Introduction

Life insurance contracts with embedded financial guarantees expose insurers to a variety of market and behavioral risks. These guarantees, common in products such as variable annuities, universal life, and participating contracts, often replicate the payoff structure of financial derivatives. Effective hedging of these insurance contracts is essential for managing risk and ensuring the insurer's solvency.

Under certain circumstances, risk managers at the insurance company can be mandated to hedge only some selected sources of risk, while leaving others unhedged. Leaving the exposure to some risk factors untouched can be justified for instance when trading the hedging instruments associated with some sources of risk is too costly; this can happen if the hedging instruments embed a substantial risk premium, if they are illiquid, or if their trading is operationally complex and requires very specialized staff. Another instance where avoiding to hedge a source of risk is desirable is when such risk generates a natural hedge within the insurer. As an example, the interest rate risk stemming from the variable annuities line of business could offset that of the conventional annuities line of business; in such case, the company might want to hedge equity risk exposure for their variable annuities potfolio, while leaving the interest rate risk untouched to avoid nullifying the natural hedge it provides. In this work, we introduce a flexible hedging framework designed specifically for the optimal hedging of a subset of risks that are targeted for mitigation, an approach we refer to as targeted hedging. Such a framework can be used for risk management in the aforementioned situations where only a subset of risk categories needs to be hedged, ensuring that hedges on targeted risk factors do not impact (or even exacerbate) exposures to unhedged risk factors through their interactions.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Actuarial Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Traditional hedging methods, such as those based on Greek sensitivities, aim to neutralize exposure to individual risk factors. These approaches rely on adjusting portfolios in response to how the guarantee's value reacts to changes in underlying risk drivers. Earlier work, such as Boyle and Hardy (1997) and Hardy (2000), applied this logic to hedge guarantee features embedded within insurance contracts. Later, Augustyniak and Boudreault (2017) considered delta-rho hedging techniques to handle fluctuations in both equity and interest rate risk factors.

While these traditional approaches are tractable and widely used in practice, they do not naturally capture the nonlinear interactions between the multiple sources of risk. A second class of methods, including those proposed by Coleman *et al*. (2006, 2007), and Kélani and Quittard-Pinon (2017), focuses on minimizing local risk and has demonstrated superior performance compared to Greeks-based hedging in several studies. These strategies perceive risk generated by the various risk factors as a whole and do not allow targeting specific risk sources. As such, these strategies are not adapted to the case where the exposure to some risk factors needs to be left untouched. Furthermore, local risk minimization provides results that are suboptimal from a multistage decision-making standpoint, as the interaction of risks over the various periods is not considered. Trottier *et al*. (2018) extend this class of methods to incorporate basis risk. A third category comprises global hedging frameworks, including dynamic programming schemes (see Schweizer 1995; Rémillard and Rubenthaler 2013; François *et al*. 2014) and, more recently, deep hedging methods (see for instance Buehler *et al*. 2019; Carbonneau 2021), which typically hedge the full insurance contract payoff. However, such strategies are not adapted to the mitigation of only a selected number of risk sources. As a result, existing approaches are sometimes misaligned with potential risk management mandates that would require targeting a selected risk category.

In this paper, we introduce a deep hedging framework that is adapted for the hedging of targeted risk sources. We build on the framework of Godin *et al*. (2023) who apply a Shapley decomposition of the variable annuity guarantee cash flows to quantify the marginal impact of each source of risk. While their paper only considers risk contributions as a risk measurement tool in the presence of conventional hedging approaches, we go further and design hedging strategies that directly aim to mitigate the targeted components. Furthermore, in contrast to Godin *et al*. (2023) who decompose local cash flows at individual time points, we apply a decomposition to cover the entire duration of the hedging period, capturing the cumulative effects of all risk factors. This allows accounting for the path-dependent dynamics of risks, while leading to a reduced computational cost and effectively managing the evolution of insurance contacts.

Recent work by François *et al*. (2025) demonstrates through numerical experiments that deep reinforement learning (RL) agents can often delay hedging portfolio adjustments until being closer to maturity, effectively maintaining low terminal risk while tolerating short-term risk. However, this behavior poses challenges for insurers, as early volatility can lead to risk limit breaches and increased capital requirements. To address this, we introduce an auxiliary neural network that penalizes extreme single-period losses during training. This encourages the agent to adopt more stable hedging strategies throughout the contract, effectively reducing short-term exposure and improving the robustness of the hedging policy.

We adopt a setup similar to that of Godin *et al*. (2023), which consists in risk mitigation of a single-premium guaranteed minimum maturity benefits (GMMB) variable annuity to evaluate the performance of our approach through numerical experiments based on Monte Carlo simulations. We also include experiments with a guaranteed minimum withdrawal benefit (GMWB). The proposed method is benchmarked against traditional delta hedging and the standard deep hedging approach. Our results show that the deep targeted hedging (DTH) method outperforms conventional deep hedging and benchmarks by fully neutralizing the targeted risk without substantially affecting other sources of risk. Unlike deep hedging that does not disentangle interrelated risks, DTH successfully isolates and mitigates the specified source of risk. Additionally, DTH shifts the distribution of the targeted risk contribution, enhancing upside potential while minimizing adverse impacts on other risk sources.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

The remainder of the paper is organized as follows. Section 2 introduces the decomposition framework used to attribute cash flow variability to distinct sources of risk and formulates the targeted hedging problem. Section 3 describes the market environment and contract features used in our simulations. Section 4 presents the numerical experiments and compares the performance of the proposed strategy against benchmark methods. Finally, Section 5 concludes the paper. Appendices contain additional technical details related to the procedure used to train neural networks, the implementation of the delta hedging benchmark, and cash flow and risk factor dynamics for the considered variable annuity model. The Python code used to replicate the numerical experiments is available at: https://github.com/cpmendoza/deep_targeted_hedging-.git.

2. The hedging problem

This section first outlines the approach to decompose the stochastic variability of cash flows into contributions from various source of risk groups. The mathematical formulation of the hedging problem integrating the decomposition and targeting specific components for risk mitigation is then presented.

2.1. Decomposing the risk exposure

Consider a $\tilde{d}$-dimensional stochastic risk factor process $X = (X^{(1)}, \ldots, X^{(\tilde{d})})^\top$ whose evolution is driven by a shock process $Y = (Y^{(1)}, \ldots, Y^{(d)})^\top$, representing the underlying sources of risk. The components of the process X may represent various financial elements, such as asset prices (e.g., stock or currency values), interest rate term structure factors, or volatility surface factors. The dynamics of X can be expressed by the following update rule, with the time-t value of the process X being given by

$$X_t = \mathcal{U}_t(X_0, Y_{1:t}),$$

where X_0 denotes the initial state of the process, $Y_{1:t} = (Y_1, \ldots, Y_t)$ is the path of shocks up to time t, and $\mathcal{U}_t : \mathbb{R}^{\tilde{d}} \times \mathbb{R}^{d\times t} \to \mathbb{R}^{\tilde{d}}$ is an update function.

Sources of risk are divided into subgroups through the partition $\mathbb{S} = \{\mathcal{S}_1, \ldots, \mathcal{S}_{\#\mathbb{S}}\}$ of $\{1, \ldots, d\}$, where $\#\mathbb{S}$ is the number of elements in the partition. Thus for any $k \in \{1, \ldots, \#\mathbb{S}\}$, $\mathcal{S}_k \subseteq \{1, \ldots, d\}$. Each group is associated with a class of sources of risk, for example, equity risk, interest rate risk, or mortality risk.

The value of insurance contracts, such as variable annuities, is often modeled as functions of underlying risk factors, which are represented by the process X. The present value at time 0 of the time-t value to the insurer of the insurance contract of interest is denoted by

$$\Pi_t = f_t(X_{0:t}),$$

where $X_{0:t}$ denotes the path of the risk factors up to time t, and f_t is a valuation function that maps this path to the corresponding contract value. The entire path of risk factors is needed, as discounting from time t to time 0 requires the full interest rate path.

The present value of the time-t contract cash flow to the insurer is

$$\widetilde{\mathrm{CF}}_t(X_{1:t}) = D_t(X_{1:t})\mathrm{CF}_t,$$

where the discount factor is

$$D_t \equiv D_t(X_{1:t}) = \exp\left(-\sum_{l=1}^{t} r(X_{l-1})\Lambda\right).$$

Here, $r(X_{l-1})$ denotes the annualized risk-free rate for the interval $[l-1, l)$, Λ represents the time step length, and the random variable CF_t represents the cash flows to the insurer, which can be expressed as a function of the current and previous-step state variables X_t and X_{t-1} (see Section 3 for more details).

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

The present value of the cumulative profit and loss (cumulative P&L) associated with the insurer's value from holding the contract until maturity is given by

$$\widetilde{\mathrm{P\&L}}_T(X_0, Y_{1:T}) \equiv f_T(X_{0:T}) - f_0(X_0) + \sum_{t=1}^{T} \widetilde{\mathrm{CF}}_t(X_{1:t})$$

$$= f_T(\mathcal{V}_T(X_0, Y_{1:T})) - f_0(X_0) + \sum_{t=1}^{T} \widetilde{\mathrm{CF}}_t(\mathcal{V}_t(X_0, Y_{1:t})),$$

where $\mathcal{V}_t(X_0, Y_{1:t}) = X_{0:t}$ for $t = 1, \ldots, T$.

Define the contribution of the passage of time to the P&L as the value of the P&L, which would occur in absence of shocks on risk factors (i.e., when $Y_{1:T} = \mathbf{0}$):

$$\Theta_T = \widetilde{\mathrm{P\&L}}_T(X_0, \mathbf{0}).$$

This leads to

$$\widetilde{\mathrm{P\&L}}_T(X_0, Y_{1:T})$$

$$= f_T(\mathcal{V}_T(X_0, Y_{1:T})) - f_T(\mathcal{V}_T(X_0, \mathbf{0})) + \sum_{t=1}^{T} [\widetilde{\mathrm{CF}}_t(\mathcal{V}_t(X_0, Y_{1:t})) - \widetilde{\mathrm{CF}}_t(\mathcal{V}_t(X_0, \mathbf{0}))] + \Theta_T. \tag{2.1}$$

Furthermore, the difference between the first two terms and the sum of cash flows differences in Equation (2.1) each can be further decomposed with a Shapley (1953)-type decomposition. For any $i = 1, \ldots, \#\mathbb{S}$, define $\mathbf{Y}_{1:t}^{\mathcal{S}_i} = (\mathbb{1}_{\mathcal{S}_i} \cdot Y_1, \ldots, \mathbb{1}_{\mathcal{S}_i} \cdot Y_t)$ with "$\cdot$" denoting the element-wise Hadamard product and $\mathbb{1}_{\mathcal{S}_i}$ being a d-dimensional vector whose elements with index in $\mathcal{S}_i$ are ones, and whose other elements are zero. Define $\mathbb{D}_k^h$ as set of possible unions of exactly h sets from $\mathbb{S}/\mathcal{S}_k$, the partition $\mathbb{S}$ from which the group of sources of risks k is removed. Then,

$$f_T(\mathcal{V}_T(X_0, Y_{1:T})) - f_T(\mathcal{V}_T(X_0, \mathbf{0}))$$

$$= \sum_{k=1}^{\#\mathbb{S}} \sum_{h=1}^{\#\mathbb{S}} \binom{\#\mathbb{S}}{h-1}^{-1} \frac{1}{\#\mathbb{S} - h + 1} \sum_{\mathcal{S} \in \mathbb{D}_k^h} \left[f_T\left(\mathcal{V}_T\left(X_0, \mathbf{Y}_{1:T}^{\mathcal{S} \cup \mathcal{S}_k}\right)\right) - f_T\left(\mathcal{V}_T\left(X_0, \mathbf{Y}_{1:T}^{\mathcal{S}}\right)\right)\right], \tag{2.2}$$

and, for $t = 1, \ldots, T$,

$$\widetilde{\mathrm{CF}}_t(\mathcal{V}_t(X_0, Y_{1:t})) - \widetilde{\mathrm{CF}}_t(\mathcal{V}_t(X_0, \mathbf{0}))$$

$$= \sum_{k=1}^{\#\mathbb{S}} \sum_{h=1}^{\#\mathbb{S}} \binom{\#\mathbb{S}}{h-1}^{-1} \frac{1}{\#\mathbb{S} - h + 1} \sum_{\mathcal{S} \in \mathbb{D}_k^h} \left[\widetilde{\mathrm{CF}}_t\left(\mathcal{V}_t\left(X_0, \mathbf{Y}_{1:t}^{\mathcal{S} \cup \mathcal{S}_k}\right)\right) - \widetilde{\mathrm{CF}}_t\left(\mathcal{V}_t\left(X_0, \mathbf{Y}_{1:t}^{\mathcal{S}}\right)\right)\right]. \tag{2.3}$$

Placing (2.2) and (2.3) into (2.1) leads to the following linear decomposition of P&L:

$$\widetilde{\mathrm{P\&L}}_T(X_0, Y_{1:T}) = \sum_{k=1}^{\#\mathbb{S}} \mathcal{C}_T^{(k,f)} + \sum_{t=1}^{T} \sum_{k=1}^{\#\mathbb{S}} \mathcal{C}_t^{(k,\mathrm{CF})} + \Theta_T,$$

where contributions from various risk factors are

$$\mathcal{C}_T^{(k,f)} = \sum_{h=1}^{\#\mathbb{S}} \binom{\#\mathbb{S}}{h-1}^{-1} \frac{1}{\#\mathbb{S} - h + 1} \sum_{\mathcal{S} \in \mathbb{D}_k^h} \left[f_T\left(\mathcal{V}_T\left(X_0, \mathbf{Y}_{1:T}^{\mathcal{S} \cup \mathcal{S}_k}\right)\right) - f_T\left(\mathcal{V}_T\left(X_0, \mathbf{Y}_{1:T}^{\mathcal{S}}\right)\right)\right],$$

$$\mathcal{C}_t^{(k,\mathrm{CF})} = \sum_{h=1}^{\#\mathbb{S}} \binom{\#\mathbb{S}}{h-1}^{-1} \frac{1}{\#\mathbb{S} - h + 1} \sum_{\mathcal{S} \in \mathbb{D}_k^h} \left[\widetilde{\mathrm{CF}}_t\left(\mathcal{V}_t\left(X_0, \mathbf{Y}_{1:t}^{\mathcal{S} \cup \mathcal{S}_k}\right)\right) - \widetilde{\mathrm{CF}}_t\left(\mathcal{V}_t\left(X_0, \mathbf{Y}_{1:t}^{\mathcal{S}}\right)\right)\right].$$

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Note that the Shapley decomposition also has the advantageous property of order invariance, which means that contributions do not depend on an order in which sources of risk are considered.

The linearity of Shapley-type decompositions with respect to various sources of risk leads to a key result, which bears similarity with that of Godin *et al*. (2023): the P&L can be linearly decomposed into contributions from different groups of sources of risk, providing a clear and additive representation of risk contributions:

$$\widetilde{\mathrm{P\&L}}_T(X_0, Y_{1:T}) = \sum_{k=1}^{\#\mathbb{S}} \mathcal{C}^{(k)} + \Theta_T, \tag{2.4}$$

where contributions from the various groups of sources of risk $k = 1, \ldots, \#\mathbb{S}$ are

$$\mathcal{C}^{(k)} = \mathcal{C}_T^{(k,f)} + \sum_{t=1}^{T} \mathcal{C}_t^{(k,\mathrm{CF})}. \tag{2.5}$$

Contributions defined in (2.5) possess the advantage of taking into consideration the impacts of co-movements in the various shocks instead of looking only at marginal shock impacts.

Our approach differs from that of Godin *et al*. (2023) in the way we define the time-t contribution. In their framework, the time-t contribution is computed by setting only the time-t value of the shock process Y to zero, while keeping all other values along the path unchanged. In contrast, we set the entire path of shocks up to time t, $Y_{1:t}$, to zero when evaluating the marginal contribution. This distinction reflects a different interpretation of risk attribution; rather than isolating the immediate impact at a single time point, our method captures the cumulative effect of all shocks up to time t. This perspective is particularly relevant in settings where the evolution of risk is path dependent, and early shocks can influence later dynamics through nonlinearities or memory effects in the update function. More importantly, our formulation leads to a reduced computational expense, as it avoids the need to simulate and evaluate counterfactual trajectories that retain portions of the original shock path while modifying only a single time point.

2.2. Hedging portfolio mechanics

We consider a hedging portfolio whose objective is to offset the portion of risk attributed to some of the sources of risk that are targeted. Our dynamic hedging strategy involves a self-financing portfolio composed of a cash account and multiple hedging instruments, whose prices depend on the risk factor process X. The portfolio is rebalanced at each time step to optimally offset net risk exposure at the contract's maturity. To accurately capture the dynamics of trading, we distinguish between the prices of hedging instruments at the beginning and end of each trading interval. Let the vector of prices at the beginning of the interval $[t, t+1)$ be denoted by $I_t^{(b)} = (I_t^{(1,b)}, \ldots, I_t^{(m,b)})$ and the vector of prices at the end of the same interval by $I_{t+1}^{(e)} = (I_{t+1}^{(1,e)}, \ldots, I_{t+1}^{(m,e)})$. Note that sometimes $I_{t+1}^{(e)} \neq I_{t+1}^{(b)}$, which occurs when the set of hedging instruments available for trading changes at time $t+1$.

The trading strategy is described by the predictable process $\delta = \{\delta_t\}_{t=1}^T$, where $\delta_t = (\delta_t^{(1)}, \ldots, \delta_t^{(m)})$ and $\delta_t^{(i)}$ denotes the number of positions on the i-th hedging instrument held over $(t-1, t]$. The time-t discounted gain or loss generated during the interval $(t-1, t]$ by the hedging portfolio is

$$\mathrm{GL}_t^{\delta}(X_0, Y_{1:t}) = \left\langle \delta_t, \left(D_t I_t^{(e)} - D_{t-1} I_{t-1}^{(b)} \right) \right\rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product.

The accumulated discounted gains generated by the hedging portfolio up to time-t are

$$G_t^{\delta}(X_0, Y_{1:t}) = \sum_{l=1}^{t} \mathrm{GL}_l(X_0, Y_{1:l}),$$

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

leading to the time-t discounted value of the self-financing portfolio,

$$V_t^{\delta}(V_0, X_0, Y_{1:t}) = V_0 + G_t^{\delta}(X_0, Y_{1:t}),$$

where V_0 is the initial hedging portfolio value.

Expanding on this, the optimal hedging action at time t, δ_{t+1}, is determined based on the available market information. Specifically, it is expressed as a function of the state variables, that is, $\delta_{t+1} = \tilde{\delta}(Z_t)$ where $\tilde{\delta}$ maps the state vector Z_t to the hedging decision. The state variables include the risk factor process X_t along with two additional variables that inform hedging decisions, namely, t and V_t^{δ}. Then, the state vector is given by $Z_t = (X_t, V_t^{\delta}, t)$.

Since prices of the hedging instruments and the risk-free asset evolve based on the risk factor process X, it is possible to linearly decompose the hedging portfolio value in a manner analogous to the P&L of the insurer's contract. By replacing the insurance contract cash flows with those of the hedging portfolio and applying the decomposition in (2.4), we denote by $\mathcal{H}_{\delta}^{(j)}$ the contribution of the source of risk group j to the hedging portfolio value. Then,

$$V_T^{\delta}(V_0, X_0, Y_{1:T}) = \sum_{j=1}^{\#\mathbb{S}} \mathcal{H}_{\delta}^{(j)} + \Theta_T^{V^{\delta}}, \tag{2.6}$$

where $\Theta_T^{V^{\delta}}$ represents the contribution of the passage of time to the hedging portfolio's cumulative P&L.

2.3. Hedging optimization

Deep reinforcement learning (deep RL) offers a powerful approach to solve hedging problems, which often take the form of a sequential decision-making problem. In this study, the approach pursued consists in approximating the optimal policy $\tilde{\delta}^*$ with a neural network $\tilde{\delta}_{\theta^{(1)}}$ with parameters $\theta^{(1)}$. Such parameters are determined by minimizing an objective function whose construction is described next.

We assume the hedger wishes to mitigate risks (i.e., offset contributions to cash flows) associated with all sources of risk within group $\mathcal{S}_{\mathcal{T}}$, with $\mathcal{T} \in \{1, \ldots, \#\mathbb{S}\}$ standing for targeted, while retaining as much as possible exposure to the remaining sources of risk. The hedging optimization thus integrates two key components that are minimized in our framework: (i) the potential net loss arising from the sources of risk in $\mathcal{S}_{\mathcal{T}}$, expressed as

$$\left\| \mathbb{1}_{\left\{\mathcal{H}_{\delta}^{(\mathcal{T})} + \mathcal{C}^{(\mathcal{T})} < 0\right\}} \left(\mathcal{H}_{\delta}^{(\mathcal{T})} + \mathcal{C}^{(\mathcal{T})}\right) \right\|_{\mathcal{L}^2}^2,$$

where the $\mathcal{L}^2$ norm of some random variable R is defined as $\|R\|_{\mathcal{L}^2} = \sqrt{\mathbb{E}[R^2]}$ and (ii) the impact of the hedging portfolio on the other sources of risk represented through

$$\sum_{\substack{k=1 \\ k \neq \mathcal{T}}}^{\#\mathbb{S}} \|\mathcal{H}_{\delta}^{(k)}\|_{\mathcal{L}^2}^2 + \|\Theta_T^{V^{\delta}}\|_{\mathcal{L}^2}^2. \tag{2.7}$$

In other words, (2.7) stipulates that the minimization attempts leaving the net hedged contribution of risk sources not targeted as close at possible to the unhedged contribution, leaving the exposure to such sources of risk untouched. In (2.7), we use the sum of individual norms for $\mathcal{H}_{\delta}^{(k)}$ and $\Theta_T^{V^{\delta}}$ to neutralize the impact of each source of risk group separately.

This entails considering the following loss function

$$L\left(\theta^{(1)}\right) = \left\| \mathbb{1}_{\left\{\mathcal{H}_{\tilde{\delta}_{\theta^{(1)}}}^{(\mathcal{T})} + \mathcal{C}^{(\mathcal{T})} < 0\right\}} \left(\mathcal{H}_{\tilde{\delta}_{\theta^{(1)}}}^{(\mathcal{T})} + \mathcal{C}^{(\mathcal{T})}\right) \right\|_{\mathcal{L}^2}^2 + \sum_{\substack{k=1 \\ k \neq \mathcal{T}}}^{\#\mathbb{S}} \left\| \mathcal{H}_{\tilde{\delta}_{\theta^{(1)}}}^{(k)} \right\|_{\mathcal{L}^2}^2 + \left\| \Theta_T^{V^{\tilde{\delta}_{\theta^{(1)}}}} \right\|_{\mathcal{L}^2}^2, \tag{2.8}$$

which is the main building block of the objective function we consider.

However, the loss function (2.8) focuses on minimization of cumulated risk at the end of the hedging period. If chosen to optimize (2.8) directly, RL strategies may inadvertently lead to high local risk,

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

defined as the risk associated with the P&L over the immediate next period. This is especially true at early stages of the hedging procedure, since the possibility to recoup early losses through subsequent gains makes such early losses less critical from the perspective of a penalty function focusing purely on cumulative losses. However, high local risk is undesirable in practice, as short-term losses must be mitigated with capital, which is costly to hold.

To address this issue, we introduce a penalty on local risk. Such penalty is applied to single-period losses. One-period-ahead net gains are defined as

$$\mathrm{NGL}_{t+1}^{\delta} = \mathrm{GL}_{t+1}^{\delta} + \widetilde{\mathrm{CF}}_{t+1} + \Pi_{t+1} - \Pi_t. \tag{2.9}$$

Such net gains account for fluctuations of both the hedging portfolio and the policy value at each rebalancing period, on top of contract cash flows. The penalization is implemented using the conditional value-at-risk (CVaR) risk measure, applied to the one-period-ahead net losses, conditional on the information available at time t. The conditional value-at-risk (CVaR) at confidence level $\alpha \in (0,1)$ for a random variable X, conditional on the information at time t, is defined as

$$\mathrm{CVaR}_\alpha(X \mid \mathcal{F}_t) = \mathbb{E}\left[X \mid X \geq \mathrm{VaR}_\alpha(X \mid \mathcal{F}_t),\ \mathcal{F}_t\right],$$

where the value-at-risk (VaR) is given by

$$\mathrm{VaR}_\alpha(X \mid \mathcal{F}_t) = \inf\left\{c \in \mathbb{R} : \mathbb{P}(X \leq c \mid \mathcal{F}_t) \geq \alpha\right\},$$

and $\mathcal{F}_t$ denotes the information available in the state space at time t. More precisely, the local risk penalization is defined by summing the conditional CVaRs on the net losses over all time periods, that is,

$$\mathrm{RC}_\alpha\left(\theta^{(1)}\right) = \sum_{t=0}^{T-1} \mathrm{CVaR}_\alpha\left(-\mathrm{NGL}_{t+1}^{\tilde{\delta}_{\theta^{(1)}}} \middle| \mathcal{F}_t\right). \tag{2.10}$$

Calculting CVaRs in (2.10) is not a trivial endeavor. The approach we pursue consists in using an auxiliary neural network that is pre-trained offline and which forecasts such local CVaRs based on simulated samples of state variables and hedging actions. Such approach is explained in Section 2.4.2 and Appendix A.2.

The objective function to be minimized in the proposed hedging optimization problem is

$$\mathcal{O}_\alpha^p(\theta^{(1)}) = L(\tilde{\delta}_{\theta^{(1)}}) + p\,\mathrm{RC}_\alpha(\tilde{\delta}_{\theta^{(1)}}), \tag{2.11}$$

where p is the weight applied to the local risk component. For illustration purposes, we consider $p = \frac{1}{T}$. However, this value could be optimized to reflect the hedger's preferences regarding the trade-off between local and global risk mitigation. The hedging agent therefore solves:

$$\min_{\theta^{(1)}} \left\{\mathcal{O}_\alpha^p\left(\theta^{(1)}\right)\right\}. \tag{2.12}$$

2.4. Neural networks architecture

We employ two separate fully connected feedforward neural networks (FFNNs) to approximate the optimal hedging strategy $\tilde{\delta}$ and the conditional CVaR. The first network serves as the RL agent, mapping the current state vector to optimal positions in the available hedging instruments. The second network estimates the conditional CVaR for the next period, taking as input both the state vector and the action selected by the RL agent. FFNNs are chosen for their flexibility and strong function approximation capabilities, as supported by the universal approximation theorem by (Hornik *et al*. 1989). This section outlines the architecture and role of each network in our framework.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

2.4.1. Hedging strategy neural network

The hedging strategy neural network with parameters $\theta^{(1)}$ is composed of four hidden layers, each with 56 neurons, and employs the ReLU activation function. Our numerical examples focus on hedging a variable annuity, which shares similarities with a short position in a European put option. This naturally leads to the use of short positions in the hedging instruments; the use of long positions would not be acceptable in our setting as these would be considered speculative actions in practice. To force the number of long positions in the hedging instruments to be negative, we design the output layer of the FFNN to map the hidden layer output, $\mathcal{Z}$, to the optimal hedging action δ_{t+1}, using a ReLU-negative activation function defined as

$$f(\mathcal{Z}, t) = \min(\mathcal{Z}, 0).$$

While we rely on an RELU-type terminal activation function, other nonpositive activations such as the negative softplus could also have been used instead.

The neural network is trained using the mini-batch stochastic gradient descent (MSGD) method. This process iteratively updates all trainable parameters of the network following the recursive equation:

$$\theta_{i+1}^{(1)} = \theta_i^{(1)} - \eta_i^{(1)} \nabla_{\theta^{(1)}} \hat{\mathcal{O}}_\alpha^p \left(\theta_i^{(1)}\right), \tag{2.13}$$

where $\theta_i^{(1)}$ represents the FFNN parameters after the i-th iteration, $\eta_i^{(1)}$ is the learning rate, $\nabla_{\theta^{(1)}}$ is the gradient with respect to $\theta^{(1)}$, and $\hat{\mathcal{O}}_\alpha^p$ represents the Monte Carlo estimate of the objective function (2.11) calculated over a mini-batch. In our experiments, the learning rate is dynamically adjusted using the Adam optimization algorithm of Kingma and Ba (2015). Further details on the MSGD implementation are provided in Appendix A.1.

The whole framework and the solution approach can indeed be framed as an RL problem. It involves taking a sequence of actions (updates to the hedging portfolio composition), each performed after observing the state of some environment that consists in market conditions. The objective is to minimize costs representing targeted factor contributions to the hedging shortfall, though a risk-aware version of the problem is considered where downside risk and tail risk of costs are minimized rather than the expected cost often considered in traditional RL algorithms (see Sutton *et al*. 2018). The solution approach (2.13) is a policy gradient, where the policy mapping state vectors into actions is refined based on a stochastic gradient descent. The algorithm can be classified as a Monte-Carlo type approach based on the terminology of Sutton *et al*. (2018) due to updates to the policy being computed only once full paths of state vectors and actions are observed and rewards on such trajectories are completed realized.

2.4.2. CVaR estimation neural network

The CVaR estimation neural network with parameters $\theta^{(2)}$ consists of three hidden layers, each with 56 neurons. It uses the ReLU activation function in the hidden layers. The output activation function is the identity function, enabling the network to predict the continuous CVaR value directly. At time t, the input to this network is the state vector (δ_{t+1}, X_t, t). The objective function for the CVaR estimation network is the mean squared error (MSE), defined as

$$\text{MSE}\left(\theta^{(2)}\right) = \mathbb{E}\left[\left(\text{CVaR}_{\theta^{(2)}}(\delta_{t+1}, X_t, t) - \text{CVaR}\left(-\text{NGL}_{t+1}^{\delta} \middle| \mathcal{F}_t\right)\right)^2\right], \tag{2.14}$$

where $\text{CVaR}\left(-\text{NGL}_{t+1}^{\delta} \middle| \mathcal{F}_t\right)$ denotes conditional CVaR of the one-period-ahead net loss under a given policy δ, and $\text{CVaR}_{\theta^{(2)}}(\delta_{t+1}, X_t, t)$ denotes the neural network's estimate of this value.

Similar to the hedging strategy network, the CVaR estimation neural network is trained using the MSGD method. The training process iteratively updates the network's parameters according to the recursive equation

$$\theta_{i+1}^{(2)} = \theta_i^{(2)} - \eta_i^{(2)} \nabla_{\theta^{(2)}} \widehat{\text{MSE}}\left(\theta_i^{(2)}\right),$$

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

where $\widehat{\text{MSE}}$ represents the Monte Carlo estimation of the objective function (2.14), calculated over a mini-batch, and the learning rate is dynamically adjusted using the Adam optimization algorithm. Further details on the MSGD implementation are provided in Appendix A.2.

3. Variable annuities and market dynamics

This section outlines the variable annuities and market dynamics used to illustrate our DTH methodology. These dynamics dictate the evolution of risk factors and establish the state space for the hedging problem. Two types of variable annuities are considered, namely a GMMB and a GMWB policy.

3.1. Variable annuities

We adopt the framework of Godin *et al*. (2023) to model the cash flows to the insurer. In this framework, policyholders deposit a premium at time $t = 0$, initializing the policy account with a value of A_0. The account value A_t evolves dynamically based on the performance of an underlying fund F_t, while fees, Fee_t, proportional to a periodic fee rate ω, are charged at the end of each period to all policyholders who are active. For the GMWB, an amount W is also withdrawn from the account at the beginning of each period. In our example, we assume a constant withdrawal and a withdrawal base of TW, implying that withdrawal guarantees are fully exhausted after T periods. The withdrawal amount is $W = 0$ for the GMMB policy.

The proportion of active policyholders at time t, denoted by ${}_t a_x$, evolves according to survival probabilities modeled as an adaptation of Lee and Carter (1992). These probabilities are driven by two main factors: random effects $\{\epsilon_t\}_{t=1}^T$, where ϵ_t are independent and identically distributed Gaussian random variables, and a systemic mortality factor $\{u_t\}_{t=1}^T$, which evolves according to a Gaussian autoregressive process of order 1 with random shocks $\{\nu_t\}_{t=1}^T$.

The framework accounts for policy lapses. Lapse penalties, Lapse penalty_t, are imposed on policyholders who terminate their policies in the early years, ensuring a sufficient level of retention within the insurer's portfolio.

The GMMB provides a benefit at maturity T to all policyholders who remain active. If the guaranteed amount G exceeds the final account value A_T, the insurer covers the shortfall, leading to a benefit payout given by:

$$\text{Benefit}_t^{GMMB} = \mathbb{1}_{\{t=T\}} {}_T a_x \max(0, G - A_T).$$

For the GMWB, the benefit at each time step for an active policyholder consists any amount in excess of the account value that needs to be paid by the insurer for the policyholder to make the full withdrawal W:

$$\text{Benefit}_t^{GMWB} = \mathbb{1}_{\{t<T\}} {}_t a_x \max(0, W - A_t).$$

The benefits, fees, and lapse penalties determine the net cash flow to the insurer at each time step. The total cash flows of the variable annuity contract are thus given by:

$$\text{CF}_t = \text{Fee}_t + \text{Lapse penalty}_t - \text{Benefit}_t.$$

The explicit formulas governing these dynamics are provided in Appendix D.

3.2. Market dynamics

The market dynamics used in our experiments are drawn from Godin *et al*. (2023) and Augustyniak *et al*. (2021) and are hereby presented.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

3.2.1. Market dynamics under the physical probability measure

We first present market dynamics considered under the physical probability measure $\mathbb{P}$. The annualized risk-free rate follows a discrete-time Gaussian three-factor model with random shocks $z_{t,i}^{(r)}$:

$$r_t = \sum_{i=1}^{3} r_t^{(i)}, \tag{3.1}$$

where each factor evolves according to

$$r_{t+1}^{(i)} = r_t^{(i)} + k_i\left(\mu_i - r_t^{(i)}\right) + \sigma_i z_{t+1,i}^{(r)}, \quad i = 1, 2, 3, \tag{3.2}$$

with $(z_{t+1,1}^{(r)}, z_{t+1,2}^{(r)}, z_{t+1,3}^{(r)})$ following a multivariate standard normal distribution whose correlation matrix is denoted by Γ. The use of multi-factor Gaussian models is standard and provides flexibility in capturing a range of yield curve shapes while ensuring model tractability.

Two equity indices are considered, with their price evolving based on exponential generalized autoregressive conditional heteroskedastic (EGARCH) dynamics. For more details about the EGARCH model, see Nelson (1991). For $j = 1, 2$:

$$\begin{aligned} R_{t+1,j}^{(S)} &= \log\frac{S_{t+1}^{(j)}}{S_t^{(j)}}, \\ R_{t+1,j}^{(S)} - r_t\Lambda &= \lambda_j^{(S)}\sqrt{h_{t,j}^{(S)}} - \frac{1}{2}h_{t,j}^{(S)} + \sqrt{h_{t,j}^{(S)}}z_{t+1,j}^{(S)}, \\ \log h_{t,j}^{(S)} &= \omega_j^{(S)} + \alpha_j^{(S)}z_{t,j}^{(S)} + \gamma_j^{(S)}\left(|z_{t,j}^{(S)}| - \frac{2}{\sqrt{2\pi}}\right) + \beta_j^{(S)}\log h_{t-1,j}^{(S)}, \end{aligned} \tag{3.3}$$

where the random shocks $(z_{t,1}^{(S)}, z_{t,2}^{(S)})$ follow a bivariate standard normal distribution, independent of the term structure innovations, with correlation ρ_{12}.

The variable annuity's underlying fund, whose time-t value is denoted by F_t, is a mixed fund that combines both equity and fixed-income assets. The returns on this fund are influenced by the previously described EGARCH processes for equity indices, as well as term structure factor variations for the fixed income portion of the fund. In addition, a volatility effect $h_t^{(F)}$ is incorporated, which evolves independently of equity and term structure shocks and is influenced by random shocks $z_t^{(F)}$ that follow a standard normal distribution. Such additional effect reflects basis risk. The mixed fund model for the variable annuity is

$$\begin{aligned} R_{t+1}^{(F)} &= \log\frac{F_{t+1}}{F_t}, \\ R_{t+1}^{(F)} - r_t\Lambda &= \xi_0 + \sum_{i=1}^{3}\xi_i(r_{t+1}^{(i)} - (1-\tilde{k}_i)r_t^{(i)}) + \sum_{j=1}^{2}\xi_j^{(S)}R_{t+1,j}^{(S)} + \sqrt{h_t^{(F)}}z_{t+1}^{(F)}, \\ \log h_t^{(F)} &= \omega^{(F)} + \alpha^{(F)}z_t^{(F)} + \gamma^{(F)}\left(|z_t^{(F)}| - \frac{2}{\sqrt{2\pi}}\right) + \beta^{(F)}\log h_{t-1}^{(F)}, \end{aligned} \tag{3.4}$$

where the innovation $z_t^{(F)}$ follows a standard normal distribution, independent of previous innovations.

3.2.2. Market dynamics under the risk-neutral probability measure

The following formulas represent the risk-neutral dynamics and parameters, with the tilde indicating risk-neutral values. The term structure factors maintain their autoregressive structure under the risk-neutral measure $\mathbb{Q}$, with dynamics for $i = 1, 2, 3$ given by:

$$r_{t+1}^{(i)} = r_t^{(i)} + \tilde{k}_i\left(\tilde{\mu}_i - r_t^{(i)}\right) + \sigma_i\tilde{z}_{t+1,i}^{(r)}, \quad \tilde{k}_i = k_i - \sigma_i\lambda_i, \quad \tilde{\mu}_i = \frac{k_i\mu_i}{\mu_i - \sigma_i\lambda_i}, \tag{3.5}$$

where λ_i represents the risk premia, for $i = 1, 2, 3$.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Table 1. *Estimates of the three-factor model.*

i	k_i	μ_i	σ_i	λ_i
1	0.00594	0.01176	0.00524	0.918
2	0.04228	−0.00043	0.00482	−5.473
3	0.02049	0.04627	0.00788	1.134

$$\Gamma = \begin{bmatrix} 1 & 0.135 & -0.787 \\ 0.135 & 1 & -0.539 \\ -0.787 & -0.539 & 1 \end{bmatrix}$$

Estimates are computed from the Canadian end-of-month yield curve data from January 1986 to October 2018.

The risk-neutral dynamics of the equity indices, for $j = 1, 2$, are given by:

$$R^{(S)}_{t+1,j} - r_t\Lambda = -\frac{1}{2}h^{(S)}_{t,j} + \sqrt{h^{(S)}_{t,j}}\tilde{z}^{(S)}_{t+1,j}, \tag{3.6}$$

$$\log h^{(S)}_{t,j} = \omega^{(S)}_j + \alpha^{(S)}_j\left(\tilde{z}^{(S)}_{t,j} - \lambda^{(S)}_j\right) + \gamma^{(S)}_j\left(|\tilde{z}^{(S)}_{t,j} - \lambda^{(S)}_j| - \frac{2}{\sqrt{2\pi}}\right) + \beta^{(S)}_j \log h^{(S)}_{t-1,j}.$$

Finally, the dynamics of the underlying fund are given by:

$$R^{(F)}_{t+1} - r_t\Lambda = -\frac{1}{2}\left(\sigma^{(F)}_t\right)^2 + \sigma^{(F)}\tilde{\epsilon}^{(F)}_{t+1}, \tag{3.7}$$

$$\log h^{(F)}_{t+1} = \omega^{(F)} + \alpha^{(F)}\left(\tilde{z}^{(F)}_{t+1} - \lambda^{(F)}_t\right) + \gamma^{(F)}\left(|\tilde{z}^{(F)}_{t+1} - \lambda^{(F)}_t| - \frac{2}{\sqrt{2\pi}}\right) + \beta^{(F)}\log h^{(F)}_t,$$

$$\left(\sigma^{(F)}_{t+1}\right)^2 = \sum_{i=1}^{3}\sum_{i'=1}^{3}\xi_i\xi_{i'}\sigma_i\sigma_{i'}\Gamma_{i,i'} + \sum_{j=1}^{2}\left(\xi^{(S)}_j\right)^2 h^{(S)}_{t+1,j}$$

$$+ 2\sum_{j=1}^{1}\sum_{k=j+1}^{2}\xi^{(S)}_j\xi^{(S)}_k\rho_{jk}\sqrt{h^{(S)}_{t+1,j}h^{(S)}_{t+1,k}} + h^{(F)}_{t+1},$$

where

$$\tilde{\epsilon}^{(F)}_{t+1} = \frac{\sum_{i=1}^{3}\xi_i\sigma_i\tilde{z}^{(r)}_{t+1,i} + \sum_{j=1}^{2}\xi^{(S)}_j\sqrt{h^{(S)}_{t,j}}\tilde{z}^{(S)}_{t+1,j} + \sqrt{h^{(F)}_t}\tilde{z}^{(F)}_{t+1}}{\sigma^{(F)}_t}, \tag{3.8}$$

$$\phi_t = \xi_0 + \sum_{i=1}^{3}\xi_i\tilde{k}_i\tilde{\mu}_i + \sum_{j=1}^{2}\xi^{(S)}_j\left(r_t\Lambda - \frac{1}{2}h^{(S)}_{t,j}\right), \tag{3.9}$$

$$\lambda^{(F)}_t = \frac{1}{\sqrt{h^{(F)}_t}}\left[\phi_t + \frac{1}{2}\left(\sigma^{(F)}_t\right)^2\right]. \tag{3.10}$$

3.2.3. Parameter estimates

In our study, we used parameters estimates from Augustyniak *et al*. (2021) for the term structure, equity indices, and underlying fund models. These are provided in the following tables: Table 1 presents parameters for the term structure model (3.2), Table 2 shows the estimated parameters for the equity indices model (3.3), and Table 3 presents parameters of the fund model (3.4).

3.3. Risk factors and state space

Based on the mortality and market dynamics outlined in Sections 3.1 and 3.2, the variable annuity P&L can be decomposed as described in Section 2.1. The sources of risk are categorized into three

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Table 2. *Estimates of the equity index model.*

Index	$\lambda_j^{(S)}$	$\omega_j^{(S)}$	$\alpha_j^{(S)}$	$\gamma_j^{(S)}$	$\beta_j^{(S)}$	ρ_{12}
S&P/TSX	0.08477	−1.0132	−0.01083	0.29438	0.84031	0.76384
S&P 500	0.12810	−1.5390	−0.16422	0.28580	0.75939	

The estimates are derived from a monthly time series of returns for the S&P/TSX Composite and S&P 500 price indices spanning February 1986 to October 2018.

Table 3. *Estimates of the fund model.*

Parameter	RBC bond fund	Assumption mixed fund
ξ_0	0.00258	0.00020
ξ_1	−6.80690	0.39095
ξ_2	−1.50116	−0.32789
ξ_3	−3.86662	−0.99664
$\xi_1^{(S)}$	0.06202	0.49773
$\xi_2^{(S)}$	0.04657	0.05913
$\omega^{(F)}$	−0.29261	−0.51325
$\alpha^{(F)}$	−0.25294	−0.06924
$\gamma^{(F)}$	0.19228	−0.20790
$\beta^{(F)}$	0.97454	0.94287

The model was estimated using monthly return data from October 2006 to October 2018 for the RBC Bond GIF Series 1 fund and from February 2002 to October 2018 for the Assumption/CI Harbour Growth & Income Fund Series A fund.

main groups: term structure innovations $(z_{t,1}^{(r)}, z_{t,2}^{(r)}, z_{t,3}^{(r)})$, equity innovations $(z_{t,1}^{(S)}, z_{t,2}^{(S)}, z_t^{(F)})$, and mortality innovations (ϵ_t, ν_t). Therefore, the vector of shocks influencing the risk factors at time t is given by

$$Y_t = \left(z_{t,1}^{(r)}, z_{t,2}^{(r)}, z_{t,3}^{(r)}, z_{t,1}^{(S)}, z_{t,2}^{(S)}, z_t^{(F)}, \epsilon_t, \nu_t\right)^\top,$$

where $\mathcal{S}_1 = \{1, 2, 3\}$, $\mathcal{S}_2 = \{4, 5, 6\}$, and $\mathcal{S}_3 = \{7, 8\}$. These shocks drive the evolution of the risk factor process, denoted by

$$X_t = \left(r_t^{(1)}, r_t^{(2)}, r_t^{(3)}, h_{t,1}^{(S)}, h_{t,2}^{(S)}, h_t^{(F)}, u_t, A_t\right)^\top,$$

which includes term structure factors, equity volatilities, the mortality factor, and the account value.

The simulation of the risk process X and of associated innovations Y is conducted using Monte Carlo simulation under the physical probability measure $\mathbb{P}$ with dynamics outlined in Appendix D and Section 3.2.1. However, to calculate the value of the GMMB and GMWB variable annuities at any time t, the process is simulated under the risk-neutral probability measure $\mathbb{Q}$. The dynamics under $\mathbb{Q}$ are provided in Section 3.2.2. An overview of the Monte Carlo simulation methodology is provided in Appendix B.

4. Numerical experiments

4.1. Market settings for numerical experiments

We consider monthly trading periods, with market dynamics, risk factor processes, and cash flow specifications following the models and parameter estimates from Section 3 and Appendix D. All parameter values, including estimation details, are borrowed from Godin *et al*. (2023).

Our analysis includes variable annuities with maturities of 20 years, that is, with $T = 240$ months. It considers two funds: the Assumption mixed fund and the RBC bond fund, whose associated parameter estimates are described in Section 3.2.3. While the Assumption fund is mixed and contains both equity

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

and fixed income investment, the RBC bond fund contains exclusively fixed income instruments. Fee rates are set to $\omega = 0.0286$ for the Assumption fund and to $\omega = 0.0206$ for the RBC bond fund. Initial risk factor values correspond to filtered estimates as of October 31, 2018, with term structure factors given by $r_0^{(1)} = -0.0690$, $r_0^{(2)} = -0.0062$, and $r_0^{(3)} = 0.0940$. Initial annualized volatilities are $\sqrt{12h_0^{(S,1)}} = 0.1412$ for the S&P/TSX, $\sqrt{12h_0^{(S,2)}} = 0.1785$ for the S&P 500, $\sqrt{12h_0^{(F)}} = 0.0090$ for the RBC fund, and $\sqrt{12h_0^{(F)}} = 0.0416$ for the Assumption fund. For simulation purposes, all fund values and the account value are initialized to one, that is, $A_0 = F_0 = S_0^{(1)} = S_0^{(2)} = 1$, with the guaranteed amount set to $G = 1$ in the GMMB case. For the GMWB case, which is not covered in Godin *et al*. (2023), we set $W = 0.0154$, which is the fair value leading to a null risk-neutral value of the guarantee (net of fees). The initial systemic mortality factor is $u_0 = -64.73868$. In all scenarios, the initial age of policyholders is set to 55 years.

The hedging instruments consist of the risk-free asset and equity futures on both market indices, with each futures being rolled over every three months. Futures positions are closed at the end of each monthly period. The prices of the two equity futures maturing at time $t+n$, observed at the beginning and end of the trading interval $[t, t+1)$, are given for $j = 1, 2$ by the following formulas:

$$I_{t,t+n}^{(j,b)} = S_t^{(j)} \tilde{P}_{t,t+n}^{(S,j)} \text{ and } I_{t,t+n}^{(j,e)} = S_{t+1}^{(j)} \tilde{P}_{t+1,t+n}^{(S,j)}, \tag{4.1}$$

$$\tilde{P}_{t,t+n}^{(S,j)} = \exp\left\{ \Lambda \sum_{i=1}^{3} m_{n,t}^{(i)} + \frac{\Lambda^2}{2} \sum_{i=1}^{3} \sum_{l=1}^{3} v_n^{(i,l)} \right\}, \tag{4.2}$$

$$m_{n,t}^{(i)} = (r_t^{(i)} - \tilde{\mu}_i) \left[\frac{1 - (1 - \tilde{k}_i)^n}{\tilde{k}_i} \right] + \tilde{\mu}_i n, \tag{4.3}$$

$$v_n^{(i,l)} = \frac{\sigma_i}{\tilde{k}_i} \frac{\sigma_l}{\tilde{k}_l} \Gamma_{i,l} \left[n - \frac{1 - (1 - \tilde{k}_i)^n}{\tilde{k}_i} - \frac{1 - (1 - \tilde{k}_l)^n}{\tilde{k}_l} + \frac{1 - (1 - \tilde{k}_i)^n (1 - \tilde{k}_l)^n}{1 - (1 - \tilde{k}_i)(1 - \tilde{k}_l)} \right]. \tag{4.4}$$

The initial hedging portfolio value is $V_0 = 0$. As such, the hedging portfolio value and the discounted gain and loss coincide: $V_t^{\delta}(V_0, X_0, Y_{1:t}) = G_t^{\delta}(X_0, Y_{1:t})$.

4.2. RL agent and CVaR network settings

This section provides the implementation details and hyperparameter choices for both the RL agent and the CVaR network.

4.2.1. RL agent parameters

RL agents are trained following the procedure outlined in Section 2.3, using a training set of 100,000 independently simulated paths. Training is conducted over 50 epochs for the GMMB and 100 epochs for the GMWB, using a mini-batch size of 1000 with learning rate parameters of 0.00005 and 0.00003, respectively. To enhance generalization, dropout regularization is applied with a probability parameter of $p = 0.5$. We do not conduct a formal hyperparameter tuning procedure, as ad hoc testing revealed that finding hyperparameters producing satisfactory performance through manual trial and error works reasonably well. The implementation is carried out in Python using TensorFlow, with neural network parameters initialized according to the method proposed by Glorot and Bengio (2010). We assess performance on a separate test set consisting of 1000 independent paths. A lower number of paths is considered to reduce the computational cost of evaluating the benchmark, which involves nested simulations that are computationally intensive and increase overall runtime. Table 4 summarizes the processing times

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Table 4. *Computation time for different components of the hedging frameworks, GMWB example.*

Task	Paths × Shapley combinations	Epochs	Computation time (min)	
			Deep targeted hedging	Delta hedging
Simulation of shock processes	401,000	NA	11	11
Produce Shapley environment (state space)	401,000 × 8	NA	59	59
Compute Δ and GMMB/GMWB price (nested simulation)	1000 × 8	NA	NA	6900
Estimation and training of the CVaR network	1000 × 8	100	20	NA
Training the deep targeted hedging agent for the GMWB	400,000 × 8	100	680	NA
Total (h)			12.8	116.2

The reported times correspond to a single experiment. Experiments and runtimes are measured on a compute node with 24 GB unified memory (shared CPU–GPU pool; capacity comparable to 24 GB system RAM but with higher effective utilization due to zero-copy access), under identical execution conditions. FFNN training uses 1 integrated GPU with 8 GPU cores. Total computation time aggregates all tasks for the corresponding hedging approach.

from a representative experimental run on the GMWB policy, reporting the runtime requirements of each computational component.

4.2.2. Conditional CVaR network parameters

The conditional CVaR network is trained on datapoints derived from 1000 paths sampled from the training set used for the RL agent. Assuming 240 time steps per path, this results in a total of 240,000 training datapoints. Training is carried out over 100 epochs with a mini-batch size of 2000 and an initial learning rate of 0.00001. As with the RL agent, dropout with a rate of $p = 0.5$ is applied to enhance generalization and mitigate overfitting. The network is implemented in Python using TensorFlow, with weights initialized according to the Glorot scheme.

4.3. Benchmarking framework

As a benchmark, we implement a delta hedging strategy that neutralizes the portfolio's delta exposure using equity futures on both underlying indices. Since the pricing function f_t has no closed-form solution and the delta depends on f_t, we estimate the required Greeks using a forward finite difference approach based on Monte Carlo simulation with a sample size of 1000. To ensure the portfolio remains fully delta-neutral, futures positions are adjusted at each rebalancing time. Detailed formulas for delta hedging are provided in Appendix C.

The comparison between our approach and the benchmark is based on the final terminal P&L of the insurer, which includes both these from the variable annuity and the hedging portfolio. It is defined as

$$\text{F-P\&L}_T = \widetilde{\text{P\&L}}_T(X_0, Y_{1:T}) + V_T(V_0, X_0, Y_{1:T}).$$

Using the hedging decomposition scheme outlined in Section 2.1, the terminal P&L is decomposed into contributions from the various risk factors as follows:

$$\text{F-P\&L}_T = \sum_{k=1}^{3} \tilde{\mathcal{C}}^{(k)} + \tilde{\Theta}, \tag{4.5}$$

where $\tilde{\mathcal{C}}^{(k)} = \mathcal{C}^{(k)} + \mathcal{H}_\delta^{(k)}$ for $k = 1, 2, 3$ represent the net contribution of the term structure, the equity, and the mortality source of risk groups, respectively. $\tilde{\Theta} = \Theta_T + \Theta_T^{V^\delta}$ represents the net contribution from the time passage risk factor. All contributions are based on the decompositions corresponding to Equations (2.4) and (2.6) for the variable annuity and hedging portfolio, respectively.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

To assess the performance of the hedging strategies, we consider three distinct measures under the physical probability measure $\mathbb{P}$: the expected final P&L, $\mathbb{E}\left[\text{F-P\&L}_T\right]$, the variance of the final P&L, $\text{Var}(\text{F-P\&L}_T)$, and the CVaR at the 95% confidence level, $\text{CVaR}_{95\%}(\text{F-P\&L}_T)$. We then perform risk allocation of these measures, which entails assessing what proportion of the risk measure values reported are caused by the various respective sources of risk groups. Obtaining the risk allocation for the various source of risk groups is essential to understand how each individual group contributes to the overall performance of the hedging strategy. The risk allocation for the k-th source of risk group is determined as follows:

- For the expected value, $\mathcal{A}_{\tilde{\mathcal{C}}^{(k)}} = \mathbb{E}\left[\tilde{\mathcal{C}}^{(k)}\right]$,
- For the variance, $\mathcal{A}_{\tilde{\mathcal{C}}^{(k)}} = \text{Cov}\left(\tilde{\mathcal{C}}^{(k)}, \sum_{i=1}^{3} \tilde{\mathcal{C}}^{(i)} + \tilde{\Theta}\right)$,
- For the CVaR, $\mathcal{A}_{\tilde{\mathcal{C}}^{(k)}} = \mathbb{E}\left[\tilde{\mathcal{C}}^{(k)} \mid \text{F-P\&L}_T > \text{VaR}_{95\%}(\text{F-P\&L}_T)\right]$.

4.4. Benchmarking hedging strategies

4.4.1. Benchmarking with delta hedging

We begin our analysis by evaluating the hedging performance of two strategies: delta hedging and targeted risk hedging with equity risk being chosen for mitigation, that is, $\mathcal{T} = 2$, which entails targeting sources of risk $(z_{t,1}^{(S)}, z_{t,2}^{(S)}, z_t^{(F)})$. To assess the robustness of our approach, we conduct this analysis on the GMMB and GMWB contracts using the two underlying funds: the Assumption mixed fund and the RBC bond fund.

We first present the results for the GMMB contract. The numerical results presented in Table 5 display the impact of hedging on risk for the two underlying funds. The various columns provide the performance metrics associated with optimal allocations for three different insurer portfolios without hedging, with delta hedging, and with targeted risk hedging.

Our numerical results in Table 5 reveal that delta hedging fails to effectively mitigate long-term risk and, in some cases, even exacerbates it. While the total variance remains nearly unchanged for the Assumption mixed fund, it increases by more than 50% for the RBC bond fund, rising from 0.00104 to 0.00162 when comparing the "No hedging" and "Delta hedging" columns. Additionally, delta hedging significantly amplifies long-term tail risk, with the CVaR rising from 0.15885 to 0.23020 for the Assumption mixed fund and from 0.09393 to 0.14004 for the RBC bond fund, as seen in the corresponding columns. This can partially be explained by the fact that the CVaR is translation invariant, and that delta hedging deteriorates total expected gains as seen in the table.

In contrast, the DTH approach successfully reduces both the variance and the CVaR, demonstrating its effectiveness in managing risk regardless of the underlying fund. When comparing the "No hedging" and "DTH" columns, total variance decreases from 0.00686 to 0.00243 for the Assumption mixed fund and from 0.00104 to 0.00086 for the RBC bond fund. Similarly, the global CVaR exposure declines from 0.15885 to 0.06014 for the Assumption mixed fund and from 0.09393 to 0.07489 for the RBC bond fund. These results highlight the superior risk management capabilities of DTH compared to delta hedging.

Additionally, our approach effectively neutralizes the equity risk component, resulting in a CVaR allocation of -0.03532 under the "DTH" column for the Assumption mixed fund. Negative values for the a CVaR allocation component indicate that such component has an offsetting effect on the CVaR, helping reducing it. In addition, the overall CVaR is reduced by 0.09871, computed as the difference between the total CVaR in the "No hedging" scenario (0.15885) and the "DTH" scenario (0.06014). This reduction closely aligns with the original equity risk allocation (0.09367), confirming the strategy's accuracy in isolating and mitigating the targeted risk source group. Furthermore, our targeted risk approach does not cause an increase in risk allocations for any of the other source of risk groups (except for the passage of time); both variance and CVaR allocations to any component other than equity risk decrease.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Table 5. *Total risk allocation of 20-year GMMB variable annuities, with and without hedging.*

Fund	Assumption mixed fund			RBC bond fund		
Hedging	No hedging	Delta hedging	DTH	No hedging	Delta hedging	DTH
Expected P&L decomposition						
Total	0.10243	0.01764	0.05351	−0.00342	−0.02922	−0.00741
Equity	−0.00356	−0.04278	0.07539	0.00276	−0.00859	0.03313
Rate	0.01682	0.00180	−0.00151	−0.01380	−0.01825	−0.00381
Mortality	0.00427	0.00454	0.00044	−0.00029	−0.00004	−0.01909
Time	0.08491	0.05407	−0.02081	0.00792	−0.00235	−0.01763
Total variance decomposition						
Total	0.00686	0.00664	0.00243	0.00104	0.00162	0.00086
Equity	0.00230	0.00225	0.00079	0.00035	0.00056	0.00030
Rate	0.00221	0.00214	0.00082	0.00035	0.00052	0.00027
Mortality	0.00235	0.00225	0.00082	0.00035	0.00054	0.00029
Time	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Total $\text{CVaR}_{95\%}$ *decomposition*						
Total	0.15885	0.23020	0.06014	0.09393	0.14004	0.07489
Equity	0.09367	0.12647	−0.03532	0.02802	0.04783	−0.00245
Rate	0.06466	0.07612	0.03897	0.04400	0.05530	0.02968
Mortality	0.08543	0.08169	0.03567	0.02983	0.03457	0.03003
Time	−0.08491	−0.05408	0.02081	−0.00792	0.00235	0.01763

Results are based on 1000 out-of-sample paths. The hedged risk is the equity risk component of a GMMB variable annuity with a maturity of $T = 240$ months, with hedging portfolios rebalanced monthly. The decomposition into contributions from the various source of risk groups is performed on the terminal portfolio value following the methodology outlined in Section 2.1. The CVaR is applied to the negative P&L to capture tail risk.

Similarly, when considering the RBC bond fund, the reduction in CVaR yields a value of 0.01904, which closely aligns with the CVaR allocation of the equity component without hedging (0.02802). At the same time, the other sources of risk remain mostly unaffected.

In addition, our hedging framework consistently yields higher expected P&L than delta hedging, highlighting its superior overall performance. While hedging strategies typically reduce profitability by the end of the hedging period, our approach avoids penalizing gains from the targeted sources of risk. As a result, the expected contribution from this targeted component tends to increase. On the other hand, delta hedging fails to effectively hedge the equity component, despite being designed to align with it. As explained in Godin *et al*. (2023), the hedge's failure can be attributed to the mismatch between the equity futures used for hedging and the actual underlying risk, a phenomenon known as basis risk. This occurs when the hedging instrument (equity futures) does not perfectly correlate with the risk exposure of the GMMB product. As a result, delta hedging tends to over-hedge the position. This over-hedging not only leads to incomplete risk mitigation, but it also dampens profitability, as the strategy relies on large positions providing high short exposure to the equity risk premium. These shortcomings are clearly reflected in the lower expected P&L reported in Table 5.

Results for the GMWB contracts are reported in Table 6. The table presents the same set of performance metrics as for the GMMB, across the two underlying funds and the three hedging strategies. The overall conclusions remain consistent: while delta hedging provides long-term risk reduction for the GMWB (except for the $\text{CVaR}_{95\%}$ metric when the underlying fund is the RBC bond fund), the equity risk component still remains substantial after hedging. In contrast, the targeted hedging framework succeeds in substantially lowering global risk exposure while neutralizing most of the targeted equity risk. This is achieved while simultaneously improving profitability in comparison to delta hedging. Thus, the superiority of the targeted hedging approach is again evident. Given that conclusions are qualitatively

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Table 6. *Total risk allocation of GMWB variable annuities, with and without hedging.*

Fund	Assumption mixed fund			RBC bond fund		
Hedging	No hedging	Delta hedging	DTH	No hedging	Delta hedging	DTH
Expected P&L decomposition						
Total	0.01780	−0.00230	0.01002	−0.00848	−0.01303	−0.00917
Equity	0.00348	−0.00595	0.03913	0.00132	−0.00065	0.01003
Rate	0.00047	−0.00024	−0.00459	−0.00106	−0.00119	−0.00549
Mortality	0.00197	0.00095	−0.00093	0.00020	0.00001	−0.00388
Time	0.01188	0.00295	−0.02358	−0.00895	−0.01121	−0.00983
Total variance decomposition						
Total	0.00167	0.00058	0.00048	0.00036	0.00032	0.00011
Equity	0.00057	0.00020	0.00006	0.00012	0.00011	0.00003
Rate	0.00056	0.00019	0.00018	0.00012	0.00011	0.00004
Mortality	0.00053	0.00019	0.00023	0.00011	0.00010	0.00004
Time	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Total $\text{CVaR}_{95\%}$ *decomposition*						
Total	0.08839	0.05682	0.02777	0.05300	0.05538	0.03841
Equity	0.03177	0.02418	−0.06212	0.01474	0.01557	−0.01983
Rate	0.03708	0.01969	0.03638	0.01464	0.01537	0.02562
Mortality	0.03142	0.01589	0.02992	0.01467	0.01323	0.02279
Time	−0.01188	−0.00294	0.02358	0.00895	0.01121	0.00983

Results are based on 1000 out-of-sample paths. The hedged risk is the equity risk component of a GMWB variable annuity, with hedging portfolios rebalanced monthly. The decomposition into contributions from the various source of risk groups is performed on the terminal portfolio value following the methodology outlined in Section 2.1. The CVaR is applied to the negative P&L to capture tail risk.

similar for the GMMB and GMWB products, we focus on the GMMB contract in the remainder of the paper.

4.4.2. Benchmarking with deep hedging

We continue our analysis by benchmarking our framework against the traditional deep hedging approach, which does not target a specific component of risk and hedges entire cash flows. For consistency, we use the same network architecture and parameters described in Section 4.2.1. However, the objective function in this baseline aims to minimize the losses of final discounted terminal P&L of the insurer, $\text{F-P\&L}_T$. Such objective function is the semi-mean squared error (SMSE), defined as

$$\|(\text{F-P\&L}_T)\mathbb{1}_{\{\text{F-P\&L}_T<0\}}\|_{\mathcal{L}^2}.$$

For illustration, we present results using the Assumption mixed fund as the underlying asset. Detailed outcomes for other underlying assets are omitted for brevity but are conceptually similar. Table 7 summarizes the optimal hedging performance metrics, with the "No hedging" case serving as a baseline.

Our numerical results demonstrate that deep hedging improves risk management by reducing allocations for both variance and CVaR compared to the no-hedging scenario. Surprisingly, these reductions are less pronounced than those achieved by our approach despite performance metrics being reported on the full $\text{F-P\&L}_T$. This can be partially explained by the use of the semi-quadratic objective function by the deep hedging agent, which does not fully align with the variance and CVaR performance metrics reported. Our results indicate that the reductions in risk allocations from our targeted risk approach are not solely attributable to the use of a multistage-optimal strategy such as deep hedging framework; rather, the use of risk targeting changes the structure of the hedging portfolio and provides incremental benefits. Specifically, as observed in the CVaR risk measure allocations, deep hedging does not fully eliminate the contribution from the equity component. While deep hedging reduces equity risk from 0.09367 (in the

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Table 7. *Total risk allocation across different RL settings for a GMMB on the assumption mixed fund.*

		RL hedging	
Hedging approach	No hedging	Deep hedging	DTH
Expected P&L decomposition			
Total	0.10243	0.09147	0.05351
Equity	−0.00356	0.00379	0.07539
Rate	0.01682	0.01532	−0.00151
Mortality	0.00427	0.00417	0.00044
Time	0.08491	0.06820	−0.02081
Total variance decomposition			
Total	0.00686	0.00510	0.00243
Equity	0.00230	0.00171	0.00079
Rate	0.00221	0.00166	0.00082
Mortality	0.00235	0.00173	0.00082
Time	0.00000	0.00000	0.00000
Total $\text{CVaR}_{95\%}$ *decomposition*			
Total	0.15885	0.10318	0.06014
Equity	0.09367	0.07378	−0.03532
Rate	0.06466	0.04192	0.03897
Mortality	0.08543	0.06068	0.03567
Time	−0.08491	−0.07321	0.02081

Results are based on 1000 out-of-sample paths. The hedged is the equity component of a GMMB variable annuity with a maturity of $T = 240$ months, with hedging portfolios rebalanced monthly. The decomposition into contributions from the various source of risk groups is performed on the terminal portfolio value following the methodology outlined in Section 2.1. The CVaR is applied to the negative P&L to capture tail risk.

"No hedging" scenario) to 0.07378 (in the "Deep hedging" scenario), our approach completely neutralizes this risk allocation, resulting in a value of −0.03532. This highlights the advantages of our targeted risk hedging framework, which effectively mitigates the targeted risk component.

Additionally, it is important to note that the SMSE of the DTH approach is 0.00032, which is slightly higher than the 0.00008 observed with deep hedging. This difference reflects the trade-off inherent in our approach, as deep hedging is specifically designed to minimize overall SMSE through its objective function.

4.5. Impact of local risk penalization on performance

In this section, we analyze the impact of local risk penalization on performance. This evaluation allows to assess the effect of incorporating the penalty term from Equation (2.10) into the objective function. For illustration, we only present results for the Assumption mixed fund underlying asset.

Table 8 presents the optimal hedging performance metrics applied to the total discounted P&L, F-P&L$_T$. The "No hedging" case serves as a baseline. The table also includes results for two RL-based hedging strategies, both labeled "RL hedging"; the column $L\left(\theta^{(1)}\right)$ presents results when a weight of $p = 0$ is assigned to the local risk component of the objective function, while the column $\mathcal{O}_\alpha^p(\theta^{(1)})$ exhibits results for the full objective function (2.11) with $p = 1/T$.

Numerical results confirm that both RL configurations effectively manage risk with respect to the cumulative P&L, as evidenced by the allocation metrics. Both the CVaR and variance are significantly lower compared to the values achieved without hedging. In particular, the risk associated with equity exposure is well managed, with allocations being close to zero for the variance and below zero for the CVaR.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Table 8. *Impact of the local penalty on total risk allocation for a GMMB on the assumption mixed fund.*

Hedging		RL hedging	
Objective function	No hedging	$L(\theta^{(1)})$	$\mathcal{O}_\alpha^p(\theta^{(1)})$
Expected P&L decomposition			
Total	0.10243	0.06992	0.05351
Equity	−0.00356	0.10531	0.07539
Rate	0.01682	−0.01186	−0.00151
Mortality	0.00427	−0.00089	0.00044
Time	0.08491	−0.02264	−0.02081
Total variance decomposition			
Total	0.00686	0.00294	0.00243
Equity	0.00230	0.00099	0.00079
Rate	0.00221	0.00100	0.00082
Mortality	0.00235	0.00095	0.00082
Time	0.00000	0.00000	0.00000
Total $\text{CVaR}_{95\%}$ *decomposition*			
Total	0.15885	0.08243	0.06014
Equity	0.09367	−0.05452	−0.03532
Rate	0.06466	0.06112	0.03897
Mortality	0.08543	0.05318	0.03567
Time	−0.08491	0.02264	0.02081

Results are based on 1000 out-of-sample paths. The hedged is the equity component of a variable annuity with a maturity of $T = 240$ months, with hedging portfolios rebalanced monthly. The decomposition into contributions from the various sources of risk groups is performed on the terminal portfolio value following the methodology outlined in Section 2.1. The CVaR is applied to the negative P&L to capture tail risk. The objective functions for the RL approach are defined as described in Section 2.3 and $L^*(\theta) = \|(\mathcal{H}_{\delta_\theta}^{(\mathcal{T})} + \mathcal{C}^{(\mathcal{T})})\mathbb{1}_{\{\mathcal{H}_{\delta_\theta}^{(\mathcal{T})}+\mathcal{C}^{(\mathcal{T})}<0\}}\|_{\mathcal{L}^2}$.

Regarding the RL configurations, we observe that the RL hedging strategy with local risk penalization, whose performance is reported in column $\mathcal{O}_\alpha^p(\theta)$, yields the lowest variance and CVaR values for the total risk encompassing all risk components. While local risk penalization is primarily aimed at improving short-term risk control, our results indicate that it also maintains or improves global risk metrics. This outcome can be attributed to the regularizing effect of local penalization, which discourages extreme fluctuations in the P&L.

Furthermore, risk allocations for the various components are all reduced by the inclusion of the penalty term in the objective function; except for the equity risk component of the CVaR, all allocation values for both the variance and CVaR are lower in the $\mathcal{O}_\alpha^p(\theta)$ than in the $L\left(\theta^{(1)}\right)$ column.

To evaluate the effect of local risk penalization throughout the hedging horizon, we analyze for various t's the CVaR of the time-t cumulative P&L defined as

$$\widetilde{\text{P\&L}}_t = \sum_{l=1}^{t} \text{NGL}_l^\delta,$$

see Section 4.6 for the definition of NGL_t^δ.

Figure 1 compares the 95%-level CVaR of $\widetilde{\text{P\&L}}_t$ for four hedging strategies: no hedging, the delta hedging benchmark, an RL agent optimized under the objective function $L(\theta)$ (which excludes local risk penalization), and an RL agent trained with the penalization term included, using the objective function $\mathcal{O}_\alpha^p(\theta)$.

Our findings show that both RL agents effectively reduce CVaR throughout the hedging period, as indicated by the yellow and green curves representing the CVaR of the RL strategies, which are lower than the blue curve for the variable annuity without hedging. However, only the agent incorporating local

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

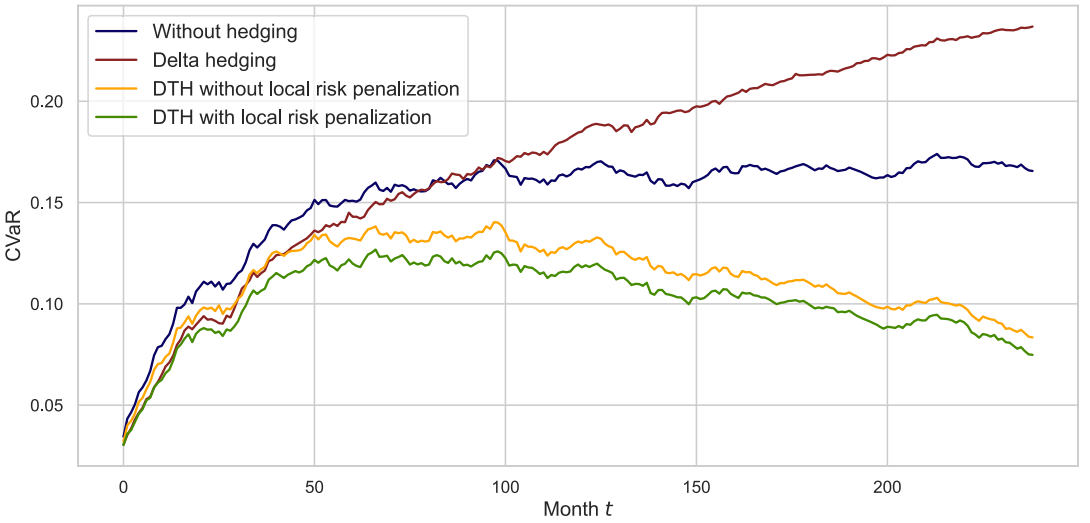


Figure 1. *CVaR of cumulative gains and losses over the hedging horizon.*
Results are based on 1000 out-of-sample paths. The hedged is the equity component of a GMMB variable annuity with a maturity of $T = 240$ months, rebalanced monthly under the assumed fund dynamics. The CVaR is computed at a 95% confidence level. In the local penalization scenario, the penalization parameter is set to $\frac{1}{T}$. The CVaR is applied to the negative P&L to capture tail risk.

risk penalization outperforms the benchmark over the whole hedging horizon, as shown by the green curve, which exhibits lower values than the red curve associated with delta hedging. This highlights the benefit of this additional control in enhancing risk management.

Together with the results presented in Table 8, these findings demonstrate that local risk penalization enhances risk control not only at the end of the hedging period but also throughout its duration.

4.6. Local contribution of risk factors

In this section, we analyze the local impact of different sources of risk groups by decomposing the period-by-period net gains NGL_t^{δ} defined in (2.9), where the net loss is decomposed in the same manner as the terminal P&L, attributing it to four distinct source of risk groups (i.e., equity, interest rates, mortality, and passage of time), as described in Equation (4.5). Specifically, the net gain decomposition for the i-th simulated path at time t is given by

$$\mathrm{NGL}_t^{(\delta,i)} = \sum_{k=1}^{3} \tilde{\mathcal{C}}_t^{(\mathrm{k},i)} + \tilde{\Theta}_t^{(i)},$$

where $\tilde{\mathcal{C}}_t^{(k,i)}$ for $k = 1, 2, 3$ represents the contribution from interest rate risk, equity risk, and mortality risk, respectively. $\tilde{\Theta}_t^{(i)}$ represents the contribution from the time passage risk factor. The purpose of examining the local net losses is to assess the contribution of individual source of risk groups and evaluate the impact of hedging on a period-by-period basis.

For illustration, we consider a GMMB variable annuity, where the underlying fund following the dynamics of the Assumption mixed fund. This case is chosen for its illustrative value, as the results remain consistent across different maturities and underlying fund choices.

Our analysis study the distribution and the average period-by-period contribution of each risk source. The average contributions are defined as

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

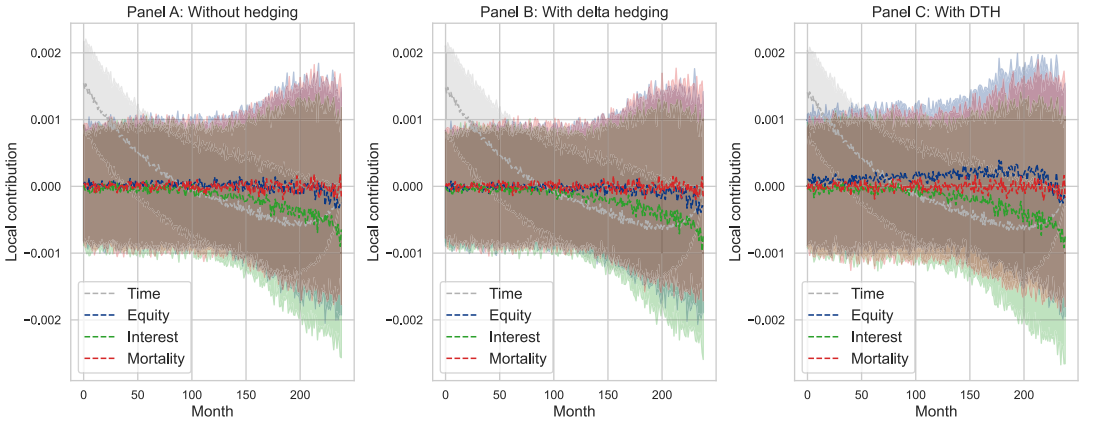


Figure 2. *Average local contribution of the risk factors to gains and losses. Results are based on 1000 out-of-sample paths. The hedged target is the equity component of a variable annuity on the Assumption fund with a maturity of $T = 240$ months. Rebalancing is performed monthly. The risk factor decomposition is computed as described in Section 2.1. The shaded confidence bands represent the interquartile range, spanning from the 25th to the 75th percentile.*

$$\widehat{\mathcal{C}}_t^{(k)} = \frac{1}{1000}\sum_{i=1}^{1000} \tilde{\mathcal{C}}_t^{(k,i)}, \text{ for } k = 1, 2, 3; \text{ and } \widehat{\Theta}_t = \frac{1}{1000}\sum_{i=1}^{1000} \tilde{\Theta}_t^{(i)}.$$

Figure 2 provides average net contributions $\widehat{\Theta}_t$ and $\widehat{\mathcal{C}}_t^{(k)}$ versus t, which are given by the full lines. Shaded areas in the figure represent the interquartile range of net contributions for each t. The Figure 2 is organized into three panels. Panel A shows the net contributions in the absence of hedging, while Panels B and C illustrate these for delta hedging and targeted risk hedging, respectively.

As shown in Figure 2, the source of risk groups exhibits varying impacts throughout the life of the contract, consistently with the findings of Godin *et al*. (2023). The average contribution of mortality risk remains close to zero over most of the hedging period, indicating a relatively minor influence on net losses compared to other sources of risk. This limited impact is partly due to the fact that our framework models only systematic mortality risk, excluding idiosyncratic mortality fluctuations that could introduce greater variability at the individual level. Nonetheless, a positive contribution is observed near maturity, which is expected given that the contract's payout structure places increasing weight on mortality outcomes as the contract nears maturity.

In contrast, the equity and interest rate factors become more volatile, with their mean exhibiting a greater deviation from zero. Since Figure 2 displays gains, lower mean values suggest that these sources of risk are contributing more to losses. Notably, the interest rate risk contribution decays early in the horizon, whereas the equity risk contribution starts to decline only a few months before maturity. This behavior can be explained by Godin *et al*. (2023), who note that, as maturity nears, only a small portion of fees remain to be collected, and the sources of risk primarily contribute to the potential for benefit appreciation. Therefore, there is limited upside risk associated with potentially higher fees charged to policyholders. Thus, shocks on risk factors are more likely to cause losses, especially due to the convexity of the value of options embedded in the contract with respect to equity rate shocks. The negative contribution from the interest rate factor emerges earlier in the horizon, due to the impact of interest rates on discount factors; impacts of interest rate movements in early stages are compounded over multiple time periods and influence the present value of a larger set of cash flows.

In terms of the effect of hedging, we observe that gains and losses under the DTH strategy primarily impact the equity component. This is evident in Panel C, where the average contributions of the non-equity source of risk groups remain consistent with those in Panels A and B. However, the equity

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

component displays a distinct pattern; its curve, in blue, shows higher values and a noticeable positive deviation from zero, indicating that DTH adjusts hedging position to increase the upside risk potential of equity risk. Such findings are consistent with the nature of targeted risk hedging and the design of its objective function, which specifically penalizes negative contributions from the equity risk component that is targeted, while aiming to minimize the impact of the hedging portfolio on contributions of other sources of risk.

5. Conclusion

We present a deep RL framework designed to optimize the hedging of specific components of risks that are targeted when performing the risk management of complex financial instruments driven by multiple risk factors. Our methodology incorporates Shapley decompositions to decompose contract cash flows into contributions from various source of risk groups, providing a detailed attribution of profits and losses to the various risk drivers. This tailored approach provides an efficient solution for managing complex hedging scenarios, particularly for instruments like variable annuities, where multiple, often interacting, risks must be addressed simultaneously.

Additionally, our approach incorporates local risk management through a novel auxiliary neural network. This auxiliary component effectively penalizes extreme values of the one-period P&L, contributing to more stable outcomes, enhancing the robustness of the hedging strategy, and reducing capital requirements to protect against short-term risk.

Through numerical experiments relying on models calibrated with real market data, we demonstrate that the proposed RL approach significantly outperforms the traditional delta hedging method. The results show marked improvements in reducing targeted risks, while leading to limited impacts on risk allocations to sources of risk that are not targeted. As a consequence, the RL-based strategy is able to substantially reduce the variance and tail risk of the overall risk exposure. Furthermore, the RL approach leads to higher profitability than delta-hedging, adequately adjusting hedge ratios to the presence of basis risk.

Our results show that DTH offers distinct advantages over conventional deep hedging by fully neutralizing the targeted risk. While deep hedging considers interactions between risks, our approach goes further by specifically isolating the sources of risk, ensuring that the targeted risk is mitigated without affecting other risk factors. This distinction is particularly valuable in scenarios where agents are tasked with hedging a specific risk while leaving others unaffected.

Furthermore, DTH is able to shift the targeted risks contribution's distribution throughout most of the hedging period, enhancing upside risk potential for such source of risk. This indicates that our approach effectively mitigates adverse impacts from the targeted source of risk while generating limited impact to contributions from other sources of risk, in accordance with the design of our objective function.

Funding statement. This work is supported by a grant from the Canadian Institute of Actuaries. Godin is funded by NSERC (RGPIN-2024-04593).

Competing interests. The authors declare none.

References

Augustyniak, M. and Boudreault, M. (2017) Mitigating interest rate risk in variable annuities: An analysis of hedging effectiveness under model risk. *North American Actuarial Journal*, **21**(4), 502–525.

Augustyniak, M., Godin, F. and Hamel, E. (2021) A mixed bond and equity fund model for the valuation of variable annuities. *ASTIN Bulletin: The Journal of the IAA*, **51**(1), 131–159.

Boyle, P.P. and Hardy, M.R. (1997) Reserving for maturity guarantees: Two approaches. *Insurance: Mathematics and Economics*, **21**(2), 113–127.

Buehler, H., Gonon, L., Teichmann, J. and Wood, B. (2019) Deep hedging. *Quantitative Finance*, **19**(8), 1271–1291.

Carbonneau, A. (2021) Deep hedging of long-term financial derivatives. *Insurance: Mathematics and Economics*, **99**, 327–340.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Coleman, T.F., Kim, Y., Li, Y. and Patron, M. (2007) Robustly hedging variable annuities with guarantees under jump and volatility risks. *Journal of Risk and Insurance*, **74**(2), 347–376.
Coleman, T.F., Li, Y. and Patron, M.-C. (2006) Hedging guarantees in variable annuities under both equity and interest rate risks. *Insurance: Mathematics and Economics*, **38**(2), 215–228.
François, P., Gauthier, G., Godin, F. and Pérez-Mendoza, C.O. (2025) Deep hedging with options using the implied volatility surface. arXiv preprint arXiv:2504.06208.
François, P., Gauthier, G. and Godin, F. (2014) Optimal hedging when the underlying asset follows a regime-switching Markov process. *European Journal of Operational Research*, **237**(1), 312–322.
Glorot, X. and Bengio, Y. (2010) Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, pp. 249–256.
Godin, F., Hamel, E., Gaillardetz, P. and Ng, E.H.-M. (2023) Risk allocation through Shapley decompositions, with applications to variable annuities. *ASTIN Bulletin: The Journal of the IAA*, **53**(2), 311–331.
Hardy, M.R. (2000) Hedging and reserving for single-premium segregated fund contracts. *North American Actuarial Journal*, **4**(2), 63–74.
Hornik, K., Stinchcombe, M. and White, H. (1989) Multilayer feedforward networks are universal approximators. *Neural Networks*, **2**(5), 359–366.
Kélani, A. and Quittard-Pinon, F. (2017). Pricing and hedging variable annuities in a Lévy market: A risk management perspective. *The Journal of Risk and Insurance*, **84**(1), 209–238.
Kingma, D.P. and Ba, J. (2015) Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*.
Lee, R.D. and Carter, L.R. (1992) Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**(419), 659–671.
Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, **59**(2), 347–370.
Rémillard, B. and Rubenthaler, S. (2013) Optimal hedging in discrete time. *Quantitative Finance*, **13**(6), 819–825.
Schweizer, M. (1995) Variance-optimal hedging in discrete time. *Mathematics of Operations Research*, **20**(1), 1–32.
Shapley, L.S. (1953) A value for n-person games. In Kuhn, H.W. and Tucker, A.W. (eds.), *Contribution to the Theory of Games*, Vol. 2. Princeton: Princeton University Press.
Sutton, R.S., Barto, A.G., et al. (2018) *Reinforcement Learning: An Introduction*, Vol. 2. Cambridge, Massachusetts: MIT Press.
Trottier, D.-A., Godin, F. and Hamel, E. (2018) Local hedging of variable annuities in the presence of basis risk. *ASTIN Bulletin: The Journal of the IAA*, **48**(2), 611–646.

A. Details for the MSGD training approach

A.1. Training approach for the hedging agent network

The MSGD method estimates the penalty function $\mathcal{O}_\alpha^p(\theta)$ using Shapley-type decompositions of the P&L for both the insurance contract and the hedging portfolio. Let $\mathbb{D}$ be the power set of $\{1, 2, 3\}$. We consider a sample $\{\{\{X_t^{(m)}(\cup_{k\in\mathcal{D}}\mathcal{S}_k)\}_{t=0}^T\}_{\mathcal{D}\in\mathbb{D}}\}_{i=1}^N$, where $X_t^{(m)}(\mathcal{S}) = \mathcal{U}_t\left(X_0^{(m)}, Y_{1:t}^{(m,\mathcal{S})}\right)$ represents the time-t risk factor values for n-th path under the set of shocked sources of risk group $\mathcal{S}$. This sample is divided into N_{batch} batches, where each batch is defined as $\mathbb{B}_i = \{\{\{X_t^{((i-1)N_{\text{batch}}+n)}(\cup_{k\in\mathcal{D}}\mathcal{S}_k)\}_{t=0}^T\}_{\mathcal{D}\in\mathbb{D}}\}_{n=1}^{B_{\text{batch}}}$, with $N = B_{\text{batch}}N_{\text{batch}}$. Each batch is then used to update the current policy parameters, $\theta_i^{(1)}$ and $\eta_i^{(1)}$.

A.2. Training approach for the CVaR estimation network

The MSGD training approach detailed in Appendix A.1 is used to train a neural network that estimates the CVaR of the net cash flow resulting from each time-t rebalancing decision, denoted by $\text{CVaR}_\alpha^{(i,n,t)}$. The estimator is modeled as a feedforward neural network, $F_{\theta^{(2)}}$, which takes as input the vector $(\delta_{t+1}^{(i,n)}, X_t^{(i,n)}, t)$ and approximates the CVaR through the mapping

$$\text{CVaR}_\alpha^{(i,n,t)} = F_{\theta^{(2)}}\left(\delta_{t+1}^{(i,n)}, X_t^{(i,n)}, t\right).$$

The neural network $F_{\theta^{(2)}}$ is trained offline following the stochastic-on-stochastic scheme described in Algorithm 2, where stochastic actions are sampled directly rather than relying on a fully optimized policy. The training process uses the same settings as in Appendix A.1, with the distinction that the sample of paths is restricted to their evolution under the influence of all risk factors, represented as $\{\{X_t^{(m)}\}_{t=0}^T\}_{m=1}^N$. Once trained, $F_{\hat{\theta}^{(2)}}$ with optimized parameters $\hat{\theta}^{(2)}$ is used as a fixed function in Algorithm 1.

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

Algorithm 1 MSGD training approach for a single epoch.

Require: Shapley decomposition of the insurance contract P&L: $\mathcal{C}^{(k)}_{(i,n)}$ for $k = 1, \ldots, \#\mathbb{S}$ and every batch-path combination (i,n).

for $i = 1$ to N_{batch} **do** ▷ Compute loss function estimation per batch

  for $n = 1$ to B_{batch} **do** ▷ Compute estimations per path

    for each $\mathcal{D} \in \mathbb{D}$**do**

      Set $\mathcal{S} = \cup_{k \in \mathcal{D}} \mathcal{S}_k$ ▷ Consider the evolution under $\mathcal{S}$

      for $t = 0$ to $T-1$ **do** ▷ Compute portfolio value under $\mathcal{S}$

        Calculate $\delta^{(i,n,\mathcal{S})}_{t+1} = \tilde{\delta}_{\theta_i}\left(Z^{(i,n,\mathcal{S})}_t\right)$ with $Z^{(i,n,\mathcal{S})}_t = \left(X^{((i-1)*N_{\text{batch}}+n)}_t(\mathcal{S}), V^{(i,n,\mathcal{S})}_t, t\right)$

        Calculate $\text{GL}^{(i,n,\mathcal{S})}_{t+1} = \left\langle \delta^{(i,n,\mathcal{S})}_{t+1}, \left(D^{(i,n,\mathcal{S})}_{t+1} I^{(e|i,n,\mathcal{S})}_{t+1} - D^{(i,n,\mathcal{S})}_t I^{(b|i,n,\mathcal{S})}_t\right)\right\rangle$

        if $\mathcal{S} = \{1, \ldots, d\}$ **then**

          Estimate $\text{CVaR}^{(i,n,t)}_\alpha = \text{CVaR}_\alpha\left(-\text{NGL}^{(i,n,\mathcal{S})}_{t+1} \middle| \mathcal{F}_t\right)$ ▷ See **Note**

        Update gains and losses $G^{(i,n,\mathcal{S})}_{t+1} = \sum_{l=1}^{t+1} \text{GL}^{(i,n,\mathcal{S})}_l$

      Compute hedging portfolio value $V^{(i,n,\mathcal{S})}_T(V_0) = \left(V_0 + G^{(i,n,\mathcal{S})}_T\right)$

    for $k = 1$ to $\#\mathbb{S}$ **do** ▷ Compute Shapley decomposition

      Calculate the contribution of the risk k, $\mathcal{H}^{(k)}_{(i,n)}$

    Compute the contribution of the passage of time, $\Theta^{V^{(i,n)}}_T$

  for $t = 0$ to $T-1$ **do** ▷ Estimate local CVaR

    $\text{CVaR}_\alpha\left(-\text{NGL}^{(i)}_{t+1} \middle| \mathcal{F}_t\right) = \frac{1}{B_{\text{batch}}} \sum_{n=1}^{B_{\text{batch}}} \text{CVaR}^{(i,n,t)}_\alpha$

  Compute local risk penalization estimate, $\widehat{\text{RC}}_\alpha\left(\tilde{\delta}_{\theta^{(1)}_i}\right) = \sum_{t=0}^{T-1} \text{CVaR}_\alpha\left(-\text{NGL}^{(i)}_{t+1} \middle| \mathcal{F}_t\right)$

  Compute estimation of loss function

  $\hat{L}\left(\tilde{\delta}_{\theta^{(1)}_i}\right) = \frac{1}{B_{\text{batch}}} \sum_{n=1}^{B_{\text{batch}}} \left(\left(\mathcal{H}^{(\mathcal{T})}_{(i,n)} + \mathcal{C}^{(\mathcal{T})}_{(i,n)}\right)^2 \mathbb{1}_{\{\mathcal{H}^{(\mathcal{T})}_{(i,n)} + \mathcal{C}^{(\mathcal{T})}_{(i,n)} < 0\}} + \sum_{\substack{k=1 \\ k \neq \mathcal{T}}}^{\#\mathbb{S}} \left(\mathcal{H}^{(k)}_{(i,n)}\right)^2 + \left(\Theta^{V^{(i,n)}}_T\right)^2\right)$

  Compute estimation of objective function $\hat{O}^p_\alpha\left(\theta^{(1)}_i\right) = \hat{L}\left(\tilde{\delta}_{\theta^{(1)}_i}\right) + p\, \widehat{\text{RC}}_\alpha\left(\tilde{\delta}_{\theta^{(1)}_i}\right)$

  $\theta^{(1)}_{i+1} \leftarrow \theta^{(1)}_i - \eta_i \nabla_{\theta^{(1)}} \hat{O}^p_\alpha(\theta^{(1)}_i)$ ▷ Gradient computed via autodiff tools

Note: The estimates of the $\text{CVaR}^{(i,n,t)}_\alpha$ are computed based on the CVaR estimation network described in Section 2.4.2.

Algorithm 2 CVaR estimation, stochastic-on-stochastic approach.

for $i = 1$ to N_{batch} **do** ▷ Compute loss function estimation per batch

  for $n = 1$ to B_{batch} **do**

    for $t = 0$ to $T-1$ **do**

      Estimate $\Pi^{(i,n)}_t$ based on the state variables $X^{((i-1)*N_{\text{batch}}+n)}_t$

      Simulate $\delta^{(i,n)}_{t+1} \sim \text{uniform}\left([l,0]^2\right)$ with $l < 0$

      Simulate M values of $I^{(i,n)}_{t+1}$, $\text{CF}^{(i,n)}_{t+1}$, $\Pi^{(i,n)}_{t+1}$

      for $m = 1$ to M **do**

        Calculate $\Pi^{(i,n,m)}_{t+1}$

        Calculate $\text{GL}^{(i,n,m)}_{t+1} = \left\langle \delta^{(i,n)}_{t+1}, \left(D^{(i,n,m)}_{t+1} I^{(e|i,n,m)}_{t+1} - D^{(i,n)}_t I^{(b|i,n)}_t\right)\right\rangle$

        Calculate $\text{NGL}^{(i,n,m)}_{t+1} = \text{GL}^{(i,n,m)}_{t+1} + \text{CF}^{(i,n,m)}_{t+1} + \Pi^{(i,n,m)}_{t+1} - \Pi^{(i,n)}_t$

        Calculate $\text{NL}^{(i,n,m)}_{t+1} = -\text{NGL}^{(i,n,m)}_{t+1}$

      Compute $\widehat{\text{CVaR}}^{(i,n,t)}_\alpha = \text{NL}^{(i,n,\lceil \alpha \cdot M \rceil)}_{t+1} + \frac{1}{(1-\alpha)M} \sum_{i=1}^{M} \max\left(\text{NL}^{(i,n,m)}_{t+1} - \text{NL}^{(i,n,\lceil \alpha \cdot M \rceil)}_{t+1}, 0\right)$

      Compute the FFNN estimate $\text{CVaR}^{(i,n,t)}_\alpha = F_{\theta^{(2)}}(\delta^{(i,n)}_{t+1}, X^{((i-1)*N_{\text{batch}}+n)}_t, t)$

      Compute the square error $\epsilon^{(i,n,t)} = \left(\text{CVaR}^{(i,n,t)}_\alpha - \widehat{\text{CVaR}}^{(i,n,t)}_\alpha\right)^2$

  Compute estimation of the mean square error $\widehat{\text{MSE}}(\theta^{(2)}_i) = \frac{1}{B_{\text{batch}} \cdot T} \sum_{n=1}^{B_{\text{batch}}} \sum_{t=0}^{T-1} \epsilon^{(i,n,t)}$

  $\theta^{(2)}_{i+1} \leftarrow \theta^{(2)}_i - \eta^{(2)}_i \nabla_{\theta^{(2)}} \widehat{\text{MSE}}\left(\theta^{(2)}_i\right)$ ▷ Gradient computed via autodiff tools

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

B. Monte Carlo simulation methodology

This appendix provides an overview of the Monte Carlo simulation methodology employed in our framework. It outlines the key steps involved in simulating risk factors.

Algorithm 3 Monte Carlo simulation.

Require: $\mathbb{D} = \mathcal{P}(\{1,2,3\})$, where $\mathcal{P}(\{1,2,3\})$ denotes the power set of $\{1,2,3\}$, representing all possible subsets of risk sources. Here, 1 corresponds to interest rate risk, 2 to equity risk, and 3 to mortality risk.

Require: Sample size N.

for $n = 1$ to N **do**

Simulate the path of random shocks $\mathbf{Y}^{(n)} = (Y_1^{(n)}, \ldots, Y_T^{(n)})$.

for each $\mathcal{D} \in \mathbb{D}$

Set $\mathcal{S} = \cup_{k \in \mathcal{D}} \mathcal{S}_k$.

Simulate the n-th path of the risk process X under the influence of shock group $\mathcal{S}$, denoted as $\{X_t^{(n)}(\mathcal{S})\}_{t=1}^T$. Here, $X_t^{(n)}(\mathcal{S}) = \mathcal{U}_t(X_0^{(n)}, \mathbf{Y}_{\mathcal{S}}^{(n)})$ represents the evolution of risk factors driven solely by the shocks $\mathcal{S}$.

Simulate the variable annuity and market components based on the risk factor process $\{X_t^{(n)}(\mathcal{S})\}_{t=1}^T$.

Store n-path simulation considering all risk factors combinations $\{\{X_t^{(n)}(\mathcal{S})\}_{t=0}^T\}_{\mathcal{S} \in \mathbb{D}}$.

Store variable annuity and market components.

C. Delta hedging details

The delta of the overall portfolio is neutralized by adjusting the futures position,

$$\Delta_{t+1}^{(S,j)} = -\frac{\Psi_t^{(\text{guar},A)} A_t \xi_j^{(S)}}{S_t^{(j)}} \cdot \frac{1}{\tilde{P}_{t,t+n}^{(S,j)}}, \tag{C.1}$$

where

$$\Psi_t^{(\text{guar},A)} = \frac{\partial f}{\partial A_t}(t, X_t). \tag{C.2}$$

Since no closed-form solution exists for the pricing function f, the Greek associated with the account value is estimated using a forward finite difference approach based on Monte Carlo simulations,

$$\frac{\partial f}{\partial A_t}(X_t) \approx \frac{\hat{f}(t, X_t + \mathbb{1}_{A_t}\epsilon) - \hat{f}(t, X_t)}{\epsilon}, \tag{C.3}$$

where $\epsilon = 0.01$ and $\mathbb{1}_A$ is an unit vector that takes the value 1 only at the position corresponding to the A_t component in X_t. Additional details on the formulas and estimation process can be found in Godin *et al*. (2023).

D. Formulation of variable annuities and market dynamics

The policy account value A_t evolves as

$$A_t = (A_{t-1} - W)\frac{F_t}{F_{t-1}}(1-\omega), \quad t = 1, \ldots, T, \tag{D.1}$$

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

where F_t is the value of the underlying fund, W is the withdrawn amount and ω is the insurer's periodic fee rate.

Fees are charged at the end of the period for all policyholders active at the beginning of the period:

$$\text{Fee}_t = {}_{t-1}a_x \, \omega \, (A_{t-1} - W)\frac{F_t}{F_{t-1}}, \quad t = 1, \ldots, T, \tag{D.2}$$

where ${}_{t-1}a_x$ is the proportion of active policyholders at time $t-1$ with the associated account value A_{t-1}.

The proportion of active policyholders at time t, denoted by ${}_ta_x$, starts with an initial value of ${}_0a_x = 1$ and evolves according to

$${}_ta_x = {}_{t-1}a_x \frac{{}_tp_x}{{}_{t-1}p_x}(1 - \mathcal{L}(m_{t-1})), \quad t = 1, \ldots, T, \tag{D.3}$$

where $\mathcal{L}(m_{t-1})$ is the lapse probability and ${}_tp_x$ denotes the probability that the policyholder survives at least t periods after time 0. For the GMMB policy, the lapse probability is function of the moneyness ratio $m_{t-1} = \frac{A_{t-1}}{G}$, whereas the lapse function assumed to be constant ($\mathcal{L}(m_t) \equiv 0.06$) for the GMWB policy. The lapse function for GMMB policies is taken from Augustyniak *et al*. (2021). The lapse rate $\mathcal{L}$ is given by

$$\mathcal{L}(m_t) = 1 - (1 - \mathcal{L}^{\text{ann}}(m_t))^{\Lambda}, \quad \mathcal{L}^{\text{ann}}(m_t) = \max\left(\gamma_1, \min\left(\gamma_2, \gamma_1 + \gamma_2\frac{(\kappa_2 - \kappa_1)}{m_t - \kappa_1}\right)\right), \tag{D.4}$$

where γ_1 and γ_2 are boundary values, and κ_1 and κ_2 adjust the lapse function shape based on moneyness. Our numerical experiments consider $\Lambda = 1/12$ and lapse parameters are set to $\gamma_1 = 0.02$, $\gamma_2 = 0.10$, $\kappa_1 = 0.4434$ and $\kappa_1 = 1.7420$.

The ratio $p_{x+t} = \frac{t+1p_x}{tp_x}$ is computed as

$$p_{x+t} = (1 - \min(1, \exp(a_{\mathcal{Y}(x+t)} + b_{\mathcal{Y}(x+t)}u_{t+1} + \epsilon_{t+1})))^{\Lambda}, \tag{D.5}$$

where $\mathcal{Y}(x+t) = \lfloor (x+t)/12 \rfloor$ and ϵ_t are i.i.d. Gaussian random variables with standard deviation given by

$$\sigma_\epsilon(\mathcal{Y}(x+t)) = \upsilon_{\epsilon,0} + \upsilon_{\epsilon,1}\mathcal{Y}(x+t) + \upsilon_{\epsilon,2}\mathcal{Y}(x+t)^2. \tag{D.6}$$

Here, a_y is the yearly base mortality rate at age y, and b_y is the yearly sensitivity of the mortality rate at age y to the systemic mortality factor u_t, which evolves according to a Gaussian autoregressive process of order 1 with a trend:

$$u_{t+1} = c + u_t + \nu_{t+1}, \tag{D.7}$$

where ν_t is a sequence of independent and identically distributed Gaussian random variables with zero mean and variance σ_ν^2.

Estimates include $\hat{c}^{\text{yr}} = -1.8325$ and $\hat{\sigma}^2_{\nu^{\text{yr}}} = 2.1957$, with monthly parameters derived as $\sigma_\epsilon^2 = \frac{1}{12}\sigma^2_{\epsilon^{\text{yr}}}$, $c = \frac{1}{12}c^{\text{yr}}$, and $\sigma_\nu^2 = \frac{\sigma^2_{\nu^{\text{yr}}}}{12}$. The standard deviation of mortality innovations is estimated via OLS regression on age, yielding $\hat{\upsilon}^{\text{ann}}_{\epsilon,0} = 0.1856$, $\hat{\upsilon}^{\text{ann}}_{\epsilon,1} = -4.050 \times 10^{-3}$, and $\hat{\upsilon}^{\text{ann}}_{\epsilon,2} = 2.769 \times 10^{-5}$. The parameters $\upsilon_{\epsilon,j}$ are then obtained by scaling the annual estimates, $\upsilon_{\epsilon,j} = \sqrt{12}\hat{\upsilon}^{\text{ann}}_{\epsilon,j}$ for $j = 0, 1, 2$. Full details are provided in Godin *et al*. (2023).

Finally, the lapse penalties are assumed to start at 7% of the account value the first year, and then to decrease by 1% every year until it reaches 0% after the seventh year. Thus they are given by

$$\text{Lapse penalty}_t = {}_{t-1}a_x \, q^{\mathcal{L}}_{x,t-1}A_t \max(0, 0.07 - 0.01\lfloor (t-1)/12 \rfloor), \quad t = 1, \ldots, T, \tag{D.8}$$

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press

where $q^{\mathcal{L}}_{x,t-1}$ is the proportion of policyholders active at $t-1$ who lapse their policy in the next period. The proportion of policyholders active at $t-1$ who lapse their policy in the next period is given by

$$q^{\mathcal{L}}_{x,t} = \frac{\lambda^{\mathcal{L}}_t}{\lambda^{\mathcal{L}}_t + \lambda^{\mathcal{M}}_t}\left(1 - \exp\left(-\left(\lambda^{\mathcal{L}}_t + \lambda^{\mathcal{M}}_t\right)\right)\right),$$

where $\lambda^{\mathcal{L}}_t = -\log\left(1 - \mathcal{L}(m_t)\right)$ and $\lambda^{\mathcal{M}}_t \equiv -\log\left(\frac{{}_{t+1}p_x}{{}_tp_x}\right)$.

For more detailed descriptions refer to Godin *et al*. (2023).

https://doi.org/10.1017/asb.2026.10084 Published online by Cambridge University Press