

Three Essays on Stochastic Processes, Reinforcement Learning, and Financial Applications

Zaniar Ahmadi

**A Thesis in The Department of
Mathematics and Statistics**

**Presented in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy (Actuarial and Financial Mathematics) at
Concordia University
Montréal, Québec, Canada**

November 2025

© Zaniar Ahmadi, 2025

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Mr. Zaniar Ahmadi**

Entitled: **Three Essays on Stochastic Processes, Reinforcement Learning,
and Financial Applications**

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Actuarial and Financial Mathematics)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

Dr. Leila Kosseim Chair of Committee

Dr. Blanka N. Horvath External Examiner

Dr. Mélina Mailhot Examiner

Dr. Yang Lu Examiner

Dr. Arusharka Sen Examiner

Dr. Frédéric Godin Thesis Co-Supervisor

Prof. Xiaowen Zhou Thesis Co-Supervisor

Approved by _____
Prof. Cody Hyndman, Chair of Department

11/10/ 2025 _____
Dr. Pascale Sicotte, Dean of Faculty

Abstract

Three Essays on Stochastic Processes, Reinforcement Learning, and Financial Applications

Zaniar Ahmadi, Ph.D.

Concordia University, 2025

In this dissertation, we undertake three research projects. In the first study, by adopting a perturbation approach, we find expressions of potential densities for refracted skew Brownian motion (skew Brownian motion with two-valued drift). As applications, we recover its transition density and study its long-time asymptotic behaviors. In addition, we also compare with previous results on transition densities for skew Brownian motions. We propose two approaches for generating quasi-random samples by approximating the cumulative distribution function and discuss their risk measurement application.

In the second study, we introduce a dynamic portfolio optimization model for an investor that is averse to both financial risk and the greenhouse gas (GHG) emissions footprint from the portfolio. The dynamic asset allocation is obtained by solving for the optimal mean-variance-emission trade-off with reinforcement learning (RL). As an application of the model, we consider the case of an investor seeking the optimal allocation across the eleven sectors of the S&P 500. We conduct an analysis for multiple investors, each with varying levels of risk and GHG aversion. Optimal RL portfolios are shown to substantially outperform equal-weighted, optimal static and myopic dynamic benchmarks by (i) properly identifying sectors offering favorable trade-off between financial and environmental performance that are in-line with the investor's preferences, (ii) adjusting the allocation dynamically to reflect changes in market and emission conditions, and (iii) considering the interaction between the various rebalancing decisions over time. Moreover, the RL optimization model demonstrates consistent performance throughout the entire investment period, not just at the final time horizon,

making it a robust strategy.

Finally, in the last study, we investigate reinforcement learning (RL) as a framework for the dynamic hedging of swaptions, contrasting its performance with traditional sensitivity-based rho-hedging. We design agents under three distinct objective functions: mean squared error, downside risk, and Conditional Value-at-Risk to capture alternative risk preferences and evaluate how these objectives shape hedging styles. Relying on a three-factor arbitrage-free dynamic Nelson-Siegel model for our simulation experiments, our findings show that most hedging effectiveness is achieved with two swaps as hedging instruments. The underlying swap consistently anchors the hedge by replicating exposure to the level factor, while a second swap complements it by mitigating residual slope and curvature risks, or by serving as a targeted tail-risk hedge under CVaR objectives. The different objectives induce distinct behaviors: the MSE agent relies on level exposures, the downside risk agent balances gain protection with selective risk-taking, and the CVaR agent prioritizes tail protection while opportunistically harvesting premia. Compared with rho-hedging, RL strategies avoid over-hedging, adapt dynamically across payoff states, reduce turnover when transaction costs are present, and remain robust under model misspecification. These results highlight RL's potential to deliver more efficient and resilient swaption hedging strategies.

Acknowledgments

I would like to express my deepest gratitude to my supervisors, **Professor Frédéric Godin** and **Professor Xiaowen Zhou**, for their invaluable guidance, encouragement, and mentorship throughout my doctoral studies. Their expertise, patience, and constant support have shaped not only this dissertation but also my growth as a researcher. I am also sincerely thankful to the members of my thesis committee for their constructive feedback and insightful comments, which helped me strengthen and refine this work.

I am deeply indebted to my **family**, and above all to my **mother**, whose love, sacrifices, and unwavering faith in me have been my constant source of strength. Her kindness and resilience continue to light my way and give meaning to every step of this path.

Finally, I would like to thank my dear **friends**, both inside and outside the university, for walking alongside me through the highs and lows of this journey. Their encouragement, laughter, and companionship have made these years not only bearable but truly memorable.

Contents

List of Figures	ix
List of Tables	xi
Introduction	1
1 A note on Refracted Skew Brownian Motion with an application	4
1.1 Introduction	4
1.2 Construction of a Skew Brownian Motion	6
1.3 Refracted Skew Brownian Motion and its exit problems	7
1.4 Potential measures of RSBM	10
1.5 Transition density of an SBM with two-valued drift	17
1.6 Long term behaviors	23
1.7 Approximation and Simulation	29
1.7.1 Approximation by mixture of truncated normal distributions	30
1.7.2 Using Deep Neural Network	31
1.7.3 Simulation Results	32
1.7.4 Application in Risk Measurement	34
1.8 Proof of Corollary 1.1	37
1.9 Proof of Proposition 1.8	40
1.10 Remarks on the joint density in Gairat and Shcherbakov [48]	43

2	Greenhouse-Gas-Emission-Aware Portfolio Optimization with Deep Reinforcement Learning	48
2.1	Introduction	48
2.2	Literature review	51
2.3	General Framework	54
2.4	Econometric model for asset returns and emissions	57
2.4.1	DCC-EGARCH model for asset returns	57
2.4.2	Greenhouse gas emissions time series model	61
2.4.3	Trade-off between financial and environmental performance	63
2.5	Numerical Results	64
2.5.1	Portfolio optimization model setting	64
2.5.2	Performance analysis	65
2.5.3	Optimal allocation analysis	70
2.5.4	Performance Comparison with Benchmarks	72
2.5.5	Regression Analysis	74
2.6	Dynamic Portfolio Optimization	76
3	Learning to Hedge Swaptions	78
3.1	Introduction	78
3.2	Literature Review	81
3.3	Market Environment and Financial Instruments	84
3.4	Hedging Framework and Replicating Portfolio Construction	89
3.4.1	Deep Hedging Using Reinforcement Learning	91
3.4.2	The Rho-Hedging Benchmark	92
3.5	Market and Model Setup	93
3.5.1	Hedging Network Architecture	94
3.5.2	Leverage Bounds	95
3.5.3	Swaption prices and their derivatives	97

3.6	Result Analysis	97
3.6.1	Hedging Performance Metrics	97
3.6.2	Hedging Performance	100
3.6.3	Residual Sensitivity Analysis	103
3.6.4	Feature Importance	107
3.6.5	Impact of Transaction Costs	110
3.6.6	Market Sensitivity Analysis	112
3.7	Zero-Coupon price	113
3.8	Factor Dynamics under the T -forward Measure	116
3.9	Model Parameter Setup	117
3.10	Swaption Pricing Network	118
3.10.1	Kolmogorov–Arnold Networks	120
3.10.2	KAN vs. MLP Performance Comparison	121

Bibliography	127
---------------------	------------

List of Figures

Figure 1.1	Histogram and CDF of generated random samples for different models with deep neural network against the exact PDF and CDF of the model	34
Figure 1.2	Histogram and CDF of generated random samples for different models with truncated Normal approximation against the exact PDF and CDF of the model . . .	35
Figure 1.3	Evaluating approximated VaR and CVaR using mixed truncated normal distribution and Monte-Carlo (MC) simulation methods for values greater than $\alpha = 0.2569$	36
Figure 2.1	Historical value of annualized Sharpe ratio versus average log-GHG intensity.	63
Figure 2.2	Heat maps of the expected value of the annualized portfolio return, its standard deviation, the associated Sharpe ratio, and the log average yearly GHG emissions of optimal portfolios for different risk and emission preference parameters ρ and γ	66
Figure 2.3	Sharpe ratio of optimal portfolios versus log-GHG emissions	68
Figure 2.4	Optimal efficient surface of portfolios in two different views.	68
Figure 2.5	Contour plots illustrating the trade-offs between expected return, risk and emissions	69
Figure 2.6	Heat maps of portfolio weights for different sectors	71
Figure 2.7	Comparison of expected wealth, wealth variance and emissions for the RL (Optimal) the equal-weighted, the optimal static (Static), and the myopic dynamic (Dynamic) portfolios for three different investors)	74

Figure 2.8	The distribution of expected return, CAR standard deviation, and log average yearly GHG emissions over the different investors (the different combinations of values for ρ and γ outlined in Section 2.5.1) for the following four strategies: RL (Optimal), optimal static (Static), myopic dynamic (Dynamic), and Equal-weighted.	75
Figure 3.1	Comparison of kernel density estimates of in-sample and out-of-sample hedging errors for RL-based strategies and dynamic rho-hedge strategies (evaluated on the out-of-sample set), across different portfolio compositions. The first row corresponds to portfolios with one swap, the second row to two swaps, and the third row to three swaps.	102
Figure 3.2	Residuals for 2-Swap Portfolio – Multiple Objectives. Shaded areas represent scaled standard deviation bands ($0.5 \times \text{std}$) for visual clarity	107
Figure 3.3	Shapley feature importance of hedge swap weights for two hedging strategies (MSE, DR, CVaR).	109
Figure 3.4	In-sample (left) and out-of-sample (right) performance of KAN and FCNN across 1000 training epochs.	122

List of Tables

Table 1.1	RSBM models with different parameters	33
Table 1.2	K-S Test Results for DNN and TNA Methods Across Different Models . . .	34
Table 2.1	Summary statistics for monthly returns of the S&P 500 sectors.	59
Table 2.2	Historical correlations between pairs of monthly returns from sectors of the S&P 500.	59
Table 2.3	DCC–EGARCH parameter estimates on monthly returns of sectors of the S&P 500.	61
Table 2.4	Summary statistics for yearly greenhouse gas intensities of each sector of the S&P 500.	62
Table 2.5	Model parameter estimates for greenhouse gas intensities of sectors of the S&P 500.	63
Table 2.6	Slopes and associated t-statistics in the regression (44) for each sector, which regresses portfolio weights of Investor 3 (moderate aversion for financial risk and emissions) on market volatility. An intercepts has been included in each regression, but it is not reported.	76
Table 3.1	Performance metrics across reinforcement learning and rho-hedging strategies using one, two, and three swaps.	101
Table 3.2	Performance metrics with transaction costs ($\delta = 10$ bps) across reinforcement learning and ρ -hedging strategies using one, two, and three swaps.	111

Table 3.3 Performance metrics under $\kappa^{\mathbb{P}}$ -shocks with transaction costs $\delta = 0.001$ (10
bps) across reinforcement learning and ρ -hedging strategies using one, two, and
three swaps. 114

Table 3.4 Performance metrics under $\theta^{\mathbb{P}}$ -shocks with transaction costs $\delta = 0.001$ (10
bps) across reinforcement learning and ρ -hedging strategies using one, two, and
three swaps. 115

Introduction

Stochastic processes and reinforcement learning (RL) stand at the intersection of probability theory, machine learning, and financial mathematics. Stochastic processes offer rigorous tools to model randomness and dynamic systems, while reinforcement learning provides a framework for sequential decision-making under uncertainty. Their integration has become increasingly important in addressing both theoretical challenges and practical applications in finance, ranging from modeling singular stochastic differential equations to managing portfolios under sustainability constraints and hedging complex derivatives.

This thesis, entitled *Three Essays on Stochastic Processes, Reinforcement Learning, and Financial Applications*, contributes to this broad literature by investigating three distinct yet connected problems. The first essay deepens the theoretical understanding of singular stochastic processes through the study of skew Brownian motion. The second essay develops a reinforcement learning-based framework for dynamic portfolio optimization that explicitly incorporates greenhouse gas (GHG) emissions, thereby contributing to sustainable finance. The third essay advances the use of reinforcement learning in risk management by applying it to the dynamic hedging of interest rate derivatives, in particular swaptions. Together, these essays demonstrate how modern probabilistic and learning techniques can be leveraged to address long-standing questions in probability while simultaneously developing innovative methods for real-world financial applications.

Skew Brownian Motion. The first essay investigates skew Brownian motion (SBM). It captures asymmetric dynamics at a boundary, making it relevant to problems involving reflecting barriers, heterogeneous media, and financial markets with frictions. While classical works provided the

foundations of SBM, many analytical properties remain challenging to establish, especially for its generalizations with drift or refracting features. In this essay, I focus on refracted skew Brownian motion (RSBM), derive its potential measures, and obtain explicit transition densities by inverting Laplace transforms. I further explore its asymptotic behavior, including stationary distributions, and compare results with previous findings on SBM with two-valued drifts. These contributions enrich the theory of stochastic differential equations and expand the mathematical toolkit available for applications in finance and applied probability.

Green Portfolio Optimization. The second essay addresses the urgent challenge of incorporating environmental considerations into investment decisions. According to the Intergovernmental Panel on Climate Change [65], average global GHG emissions in the last decade reached record levels, prompting regulators and investors to promote sustainable and responsible finance. While ESG (Environmental, Social, and Governance) scores are widely used, they suffer from subjectivity, inconsistency, and vulnerability to greenwashing. To overcome these issues, I propose a portfolio optimization framework that directly incorporates measurable GHG emissions intensities into the decision-making process. Building on mean-variance portfolio theory, I develop a dynamic mean-variance-environment (mean-var-E) model, solved through reinforcement learning. This approach accounts for financial returns, risk, and GHG emissions simultaneously, adapting to evolving market conditions in a multi-period setting. Empirical analysis on S&P 500 sector data demonstrates that the proposed method outperforms static and myopic benchmarks by achieving higher expected returns, lower risks, and reduced emissions. This essay contributes to the growing field of sustainable finance by providing a robust, adaptive, and quantitative framework for green portfolio management.

Swaption Hedging with Reinforcement Learning. The third essay turns to risk management of interest rate derivatives, focusing on swaptions, which play a central role in global fixed-income markets. Swaptions are vital for hedging interest rate volatility and convexity risk. Traditional hedging strategies rely on sensitivity-based approaches within parametric models such as Gaussian

multi-factor models or the LIBOR Market Model. While tractable, these methods are inherently local, often failing to account for path-dependence, compounding risk premia, and higher-order exposures. In this work, I reformulate swaption hedging as a sequential decision-making problem and apply deep reinforcement learning to develop adaptive, multi-stage hedging strategies. Agents are trained under alternative objective functions, including mean squared error, downside risk, and Conditional Value-at-Risk (CVaR), highlighting how RL can flexibly encode different risk preferences. Numerical experiments show that RL-based hedging achieves superior performance, lower transaction costs, and robustness to model misspecification compared to conventional approaches. This essay contributes to the literature on financial risk management by demonstrating the potential of reinforcement learning as a flexible and economically meaningful framework for hedging interest rate derivatives.

Contributions of the Thesis. Taken together, these three essays highlight the versatility of stochastic processes and reinforcement learning in addressing both theoretical and applied challenges. The first essay advances the mathematical theory of stochastic differential equations, with implications for applied probability and finance. The second essay develops a reinforcement learning–based portfolio optimization framework that integrates financial returns, risk, and environmental sustainability. The third essay introduces a reinforcement learning paradigm for dynamic hedging of swaptions, improving upon traditional sensitivity-based methods. Beyond their individual contributions, the essays collectively illustrate how probabilistic modeling and machine learning can be combined to generate both theoretical insights and practical solutions in financial mathematics. The results of this thesis thus contribute to advancing both the academic understanding of stochastic processes and the development of innovative, real-world applications in finance.

Chapter 1

A note on Refracted Skew Brownian Motion with an application

1.1 Introduction

Skew Brownian motion (SBM for short) was first introduced by Itô and McKean in 1965 in Section two of their book [66]. In 1978 Walsh [110] proposed a path construction of skew Brownian motion and found an expression for the transition density of skew Brownian motion. Harrison and Shepp [58] showed that skew Brownian motion is the weak limit of a symmetric random walk on the integers with modified behavior at the origin, and also proved that it is the unique strong solution to a stochastic differential equation driven by Brownian motion together with a singular drift of its local time at 0. In 1984, Le Gall [77] further established the existence and pathwise uniqueness of strong solutions to a more general stochastic differential equation with a general diffusion coefficient involving local times at different levels. Appuhamillage et al. in [8] for the first time analyzed the skew Brownian motion with one drift and provided the joint distribution of key functionals of skew Brownian motion. We refer to Lejay in [80] for a review of earlier work related to skew Brownian motion.

Skew Brownian motion is a versatile concept with applications spanning diverse domains.

Notably, in mathematical finance, it serves as a fundamental tool for option pricing, as evidenced by studies such as those by Corns and Satchell [32], Wang et al. [112], Gairat and Shcherbakov [48], and Zhu and He [117]. Additionally, it finds applications in the interest rate and intensity modeling, with notable contributions from Jarrow et al. [69], Duffie and Singleton [40], and Decamps et al. [34]. Calibrating skew Brownian motion parameters has been the subject of intensive research. Bardou and Martinez [14] introduced a statistical estimator for calibrating the coefficients of a skewed process in the case in which it is reflected over a bounded interval. Bai et al. [12] advocated for a Bayesian approach to estimate the skew coefficient and skew level of the skew Ornstein-Uhlenbeck (OU) process, and Lajay and Mazzonetto [81] proposed the use of Maximum Likelihood Estimators to calibrate the parameters of a skew Brownian motion. Furthermore, skew Brownian motion is instrumental in addressing dispersion in heterogeneous media. Noteworthy contributions include works by Freidlin and Sheu [47], Ramirez [100], and Appuhamillage et al. [8]. Moreover, it proves invaluable in modeling diffusion problems with spatially discontinuous diffusion coefficients, as highlighted in studies by Zhang [116] and Ascione et al. [10].

It is an interesting problem to identify probability laws for solutions to singular SDEs. The trivariate density of Brownian Motion and its local and occupation time were obtained in Karatzas and Shreve [71] and applied to problems of stochastic control. An expression for transition density of skew Brownian motion was first found by Walsh [110]. Berezin and Zayats [15] found an expression of transition density for a skew Brownian motion with alternating drift (or skew Caughey-Dieness process as mentioned in their paper) in [15]. Gairat and Schenrbakov [48] used an approximation method for obtaining the joint distribution of both Brownian motion and skew Brownian motion with two-valued drift and their functionals. To enhance precision, the researchers approximate skew Brownian motion with two-valued drift using a discrete Markov chain. Except at skew level zero, this Markov chain evolves in the same way as a simple random walk with unit jumps. Notably, the authors employ a complex and lengthy methodology to determine transition probabilities. However, significant details about this method are excluded from their explanation. The same authors obtained the probability density of skew Brownian motion with alternating drifts in Gairat and Schenrbakov

[49]. In 2023, Borodin [20] found the potential measure of skew Brownian motion with two-valued drift using the Laplace transform method to solve the associated partial differential equation, and calculated the distributions of integral its functionals with respect to spatial variable of local time.

In this paper, we find expressions of potential measures for refracted skew Brownian motion (RSBM) and then obtain an expression for transition density by inverting the Laplace transform. Applying the above results we further discuss the asymptotic behaviors of the skew Brownian motion with two-valued drift. To this end we adopt a new perturbation approach of Chen et al. [28] that is different from those in [20] and [48]. More precisely, we take use of the solution of RSBM in Gao et al. [50] and the previous results on the potential measure of Brownian motion with drift. We then discuss the long-term behaviors of the skew Brownian motion such as the stationary density. We also compare our results with previous similar results on Skew Brownian motions.

The rest of the paper is arranged as follows. In Section 1.3 we introduce the skew Brownian motion with two-valued drift and its exit problems. In Section 1.4 we derive expressions for the potential measures applying solutions to the exit problems together with a perturbation approach. In Section 1.5 we invert the Laplace transform to find explicit formulas for the transition densities. In Section 1.6 we discuss the limiting behaviors of the refracted skew-Brownian motion. Finally, In section 1.7 we propose two approaches to estimate the cumulative distribution function to generate quasi-random samples of SBM. These approximations are also used to estimate two conventional risk metrics, Value at Risk (VaR) and Conditional Value at Risk (CVaR).

1.2 Construction of a Skew Brownian Motion

Here, we introduce a construction of skew Brownian motion from a skew random walk. Divide the half-line $[0, \infty)$ to tiny subintervals of length δ . Each subinterval corresponds to a time slot of length δ . Thus, the intervals are $(0, \delta], (\delta, 2\delta], (2\delta, 3\delta], \dots$. More generally, the k th interval is $((k-1)\delta, k\delta]$. We define a skew random walk $S = (S_i)_{i \geq 0}$ as starting from 0 with probability

transition:

$$\mathbb{P}(S_{i+1} = m + 1 | S_i = m) = \begin{cases} \frac{1+\beta}{2}, & m = 0, \\ \frac{1}{2}, & \text{otherwise,} \end{cases}$$

$$\mathbb{P}(S_{i+1} = m - 1 | S_i = m) = \begin{cases} \frac{1-\beta}{2}, & m = 0, \\ \frac{1}{2}, & \text{otherwise,} \end{cases}$$

where $\beta \in (-1, 1)$ is skew parameter. In a special case $\beta = 0$, S is a simple symmetric random walk on $\mathbb{Z}$. We set for all $t \geq 0$ and any integer n ,

$$X_t^n = \frac{1}{n} S_{\lfloor n^2 t \rfloor} + \frac{n^2 t - \lfloor n^2 t \rfloor}{n} (S_{1+\lfloor n^2 t \rfloor} - S_{\lfloor n^2 t \rfloor}).$$

Harrison and Shepp [58], and then Le Gall [78] and Cherny et al. [29] show that the sequences of $(X^n)_{n \in \mathbb{N}}$ converge in distribution in the space of continuous functions to skew Bronian motion X ,

$$X_t = x + B_t + \beta L_t^0(X),$$

where B_t is standard Brownian motion and $L_t^0(X)$ is the local time at 0 of X , meaning that:

$$L_t^0(X) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \text{Leb}(s \leq t : |X_s| \leq \epsilon).$$

Harrison and Shepp [58] show that the above stochastic equation has a unique solution.

1.3 Refracted Skew Brownian Motion and its exit problems

In this section, we consider a refracted skew Brownian motion, $X = (X_t)_{t \geq 0}$, defined on a filtered probability space $(\mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, which is the pathwise unique solution to SDE

$$X_t = X_0 + B_t + \beta L^X(t, a) + \mu_+ \int_0^t \mathbb{1}_{\{X_s > a\}} ds + \mu_- \int_0^t \mathbb{1}_{\{X_s < a\}} ds, \quad t \geq 0 \quad (1)$$

where $X_0, \mu_+, \mu_- \in \mathbb{R}$ and a is skew level, and $\beta \in (-1, 1)$, B denotes a standard Brownian motion and

$$L^X(t, a) := \lim_{\epsilon \rightarrow 0^+} \frac{1}{2\epsilon} \text{Leb}(s \leq t : |X_s - a| \leq \epsilon),$$

denotes the symmetric local time process for X at level a up to time $t \in \mathbb{R}_+$, see e.g. Le Gall [77] for existence and uniqueness of solution to (1).

Remark 1. Note that for $\sigma > 0$ and $Y_0, \mu_-, \mu_+ \in \mathbb{R}$, Y solves SDE

$$Y_t = Y_0 + \sigma B_t + \beta \sigma L^Y(t, \sigma a) + \sigma \mu_+ \int_0^t \mathbb{1}_{\{Y_s > \sigma a\}} ds + \sigma \mu_- \int_0^t \mathbb{1}_{\{Y_s < \sigma a\}} ds$$

if and only if $X_t := Y_t/\sigma$ solves SDE

$$X_t = Y_0/\sigma + B_t + \beta L^X(t, a) + \mu_+ \int_0^t \mathbb{1}_{\{X_s > a\}} ds + \mu_- \int_0^t \mathbb{1}_{\{X_s < a\}} ds.$$

For a boundary $y \in \mathbb{R}$, define the first hitting time of the boundary y for the stochastic process X_t to be

$$\tau_y := \inf\{t \geq 0 : X_t = y\},$$

with the convention $\inf \emptyset = \infty$. We denote the Laplace transform of τ_y as $\mathbb{E}_x[e^{-q\tau_y}]$ for a fixed $q \geq 0$. For any interval $[y, z]$, where $x, y, z \in \mathbb{R}$ and $y \leq x < z$, we denote the first hitting time for the lower barrier y and upper barrier z for the stochastic process $\{X_t\}_{t \geq 0}$ as τ_y and τ_z , respectively. Then define the first exit time of the interval (y, z) by

$$\tau_{y,z} := \tau_y \wedge \tau_z = \min\{\tau_y, \tau_z\}.$$

Moreover, the Laplace transform of τ_y for $\tau_y < \tau_z$ is expressed as $\mathbb{E}_x[e^{-q\tau_y}; \tau_y < \tau_z]$. The Laplace transform of τ_z for $\tau_z < \tau_y$ is expressed as $\mathbb{E}_x[e^{-q\tau_z}; \tau_z < \tau_y]$. It can be interpreted as the Laplace transform of the time of ruin before reaching the lower barrier y and plays a key role in the problems of negative duration and occupation time.

Note that SDE (1) has a pathwise unique solution. To obtain the Laplace transform of some exit times, we first need to find the general solutions $g_{1,q}(x)$ and $g_{2,q}(x)$ in $C(\mathbb{R}) \cap C^2(\mathbb{R} \setminus \{a\})$ to the following equation:

$$\frac{1}{2}g''(x) + \mu_+ \mathbf{1}_{\{x>a\}}g'(x) + \mu_- \mathbf{1}_{\{x<a\}}g'(x) = qg(x),$$

with $(1 + \beta)g'(a^+) = (1 - \beta)g'(a^-)$ for $\beta \in (-1, 1)$ and $q \geq 0$.

Let $\delta^- = \sqrt{\mu_-^2 + 2q}$, and $\rho_1^- = -(\mu_- + \delta^-)$ and $\rho_2^- = -\mu_- + \delta^-$ be the root of equation $\frac{1}{2}x^2 + \mu_-x - q = 0$, where $\rho_1^- < 0 < \rho_2^-$. Beside that, let $\delta^+ = \sqrt{\mu_+^2 + 2q}$, and $\rho_1^+ = -(\mu_+ + \delta^+)$ and $\rho_2^+ = -\mu_+ + \delta^+$ be the root of equation $\frac{1}{2}x^2 + \mu_+x - q = 0$, where $\rho_1^+ < 0 < \rho_2^+$. Let

$$g_{1,q}(x) = e^{\rho_1^+(x-a)} \mathbf{1}_{\{x>a\}} + \left(c_1(q)e^{\rho_2^-(x-a)} + (1 - c_1(q))e^{\rho_1^-(x-a)} \right) \mathbf{1}_{\{x \leq a\}} \quad (2)$$

and

$$g_{2,q}(x) = \left((1 - c_2(q))e^{\rho_2^+(x-a)} + c_2(q)e^{\rho_1^+(x-a)} \right) \mathbf{1}_{\{x>a\}} + e^{\rho_2^-(x-a)} \mathbf{1}_{\{x \leq a\}}, \quad (3)$$

where

$$c_1(q) = \frac{(1 + \beta)\rho_1^+ - (1 - \beta)\rho_1^-}{(1 - \beta)(\rho_2^- - \rho_1^-)}$$

and

$$c_2(q) = \frac{(1 + \beta)\rho_2^+ - (1 - \beta)\rho_2^-}{(1 + \beta)(\rho_2^+ - \rho_1^+)}.$$

The following two theorems are presented by Gao et al. [50] and serve as the foundation for derivations in the preceding sections.

Theorem 1.1. *For $x, y, z \in \mathbb{R}$ and $y \leq x < z$, we have*

$$\mathbb{E}_x[e^{-q\tau_z}, \tau_z < \tau_y] = \frac{w(x, y)}{w(z, y)} \quad (4)$$

and

$$\mathbb{E}_x[e^{-q\tau_y}; \tau_y < \tau_z] = \frac{w(x, z)}{w(y, z)}, \quad (5)$$

where

$$w(x, y) := g_{2,q}(x)g_{1,q}(y) - g_{1,q}(x)g_{2,q}(y).$$

Theorem 1.2. For $x, r \in \mathbb{R}$ and $x \geq r$,

$$\mathbb{E}_x[e^{-q\tau_r}] = e^{\rho_1^+(x-r)} \mathbf{1}_{\{r > a\}} + \frac{g_{1,q}(x)}{c_1(q)e^{\rho_2^-(r-a)} + (1 - c_1(q))e^{\rho_1^-(r-a)}} \mathbf{1}_{\{r \leq a\}} \quad (6)$$

and for $x < r$,

$$\mathbb{E}_x[e^{-q\tau_r}] = e^{\rho_2^-(x-r)} \mathbf{1}_{\{r \leq a\}} + \frac{g_{2,q}(x)}{(1 - c_2(q))e^{\rho_2^+(r-a)} + c_2(q)e^{\rho_1^+(r-a)}} \mathbf{1}_{\{r > a\}}. \quad (7)$$

1.4 Potential measures of RSBM

Throughout this paper, we assume e_q has an exponential distribution with intensity q that is independent of X_t .

Using the results of Theorem 1.1 and Theorem 1.2, the memoryless property of exponential distribution, and the Markov property of skew Brownian motion, we can derive the following theorem.

Theorem 1.3. The potential measure of the refracted skew Brownian motion X_t with skew level $a = 0$ and starting point $x = 0$ is given by:

$$\mathbb{P}(X_{e_q} \in dy) = \begin{cases} \frac{2(1+\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} e^{(\mu_+ - \sqrt{\mu_+^2 + 2q})y} dy, & y > 0, \\ \frac{2(1-\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} e^{(\mu_- + \sqrt{\mu_-^2 + 2q})y} dy, & y < 0. \end{cases} \quad (8)$$

Proof. Let $X_t^1 = \mu_+ t + B_t$ be a Brownian motion and let τ_x^1 be the first hitting time of process X_t^1 .

For $0 < \epsilon < y$, by the Markov property and the memoryless property,

$$\begin{aligned}
\mathbb{P}(X_{e_q} \in dy) &= \mathbb{P}(\tau_\epsilon < e_q, X_{e_q} \in dy) \\
&= \mathbb{P}(\tau_\epsilon < e_q) \mathbb{P}_\epsilon(X_{e_q} \in dy) \\
&= \mathbb{E} e^{-q\tau_\epsilon} \left(\mathbb{P}_\epsilon(\tau_0 < e_q, X_{e_q} \in dy) + \mathbb{P}_\epsilon(e_q < \tau_0, X_{e_q} \in dy) \right) \\
&= \mathbb{E} e^{-q\tau_\epsilon} \left(\mathbb{E}_\epsilon e^{-q\tau_0} \mathbb{P}(X_{e_q} \in dy) + \mathbb{P}_\epsilon(e_q < \tau_0^1, X_{e_q}^1 \in dy) \right).
\end{aligned} \tag{9}$$

Solving the above equation we have

$$\begin{aligned}
\int_0^\infty q e^{-qt} \mathbb{P}(X_t \in dy) dt &= \mathbb{P}(X_{e_q} \in dy) \\
&= \lim_{\epsilon \rightarrow 0^+} \frac{\mathbb{E} e^{-q\tau_\epsilon} \mathbb{P}_\epsilon(e_q < \tau_0^1, X_{e_q}^1 \in dy)}{1 - \mathbb{E} e^{-q\tau_\epsilon} \mathbb{E}_\epsilon e^{-q\tau_0}} \\
&= \lim_{\epsilon \rightarrow 0^+} \frac{\mathbb{E} e^{-q\tau_\epsilon} \left(\mathbb{P}_\epsilon(X_{e_q}^1 \in dy) - \mathbb{P}_\epsilon(e_q > \tau_0^1) \mathbb{P}(X_{e_q}^1 \in dy) \right)}{1 - \mathbb{E} e^{-q\tau_\epsilon} \mathbb{E}_\epsilon e^{-q\tau_0}} \\
&= \lim_{\epsilon \rightarrow 0^+} \frac{\mathbb{E} e^{-q\tau_\epsilon} \left(\mathbb{P}_\epsilon(X_{e_q}^1 \in dy) - \mathbb{E}_\epsilon e^{-q\tau_0} \mathbb{P}(X_{e_q}^1 \in dy) \right)}{1 - \mathbb{E} e^{-q\tau_\epsilon} \mathbb{E}_\epsilon e^{-q\tau_0}} \\
&= \lim_{\epsilon \rightarrow 0^+} \frac{\mathbb{P}_\epsilon(X_{e_q}^1 \in dy) - e^{\rho_1^+ \epsilon} \mathbb{P}(X_{e_q}^1 \in dy)}{(1 - c_2(q)) e^{\rho_2^+ \epsilon} + (c_2(q) - 1) e^{\rho_1^+ \epsilon}} \quad \text{\textit{l'Hôpital's rule}} \tag{10} \\
&= \lim_{\epsilon \rightarrow 0^+} \frac{\frac{d}{d\epsilon} \mathbb{P}_\epsilon(X_{e_q}^1 \in dy) - \rho_1^+ e^{\rho_1^+ \epsilon} \mathbb{P}(X_{e_q}^1 \in dy)}{(1 - c_2(q)) \rho_2^+ e^{\rho_2^+ \epsilon} + (c_2(q) - 1) \rho_1^+ e^{\rho_1^+ \epsilon}} \\
&= \lim_{\epsilon \rightarrow 0^+} \frac{\frac{q}{\delta^+} (-\mu_+ + \delta^+) e^{(\mu_+ - \delta^+)(y - \epsilon)} - \rho_1^+ e^{\rho_1^+ \epsilon} \frac{q}{\delta^+} e^{(\mu_+ - \delta^+)y}}{(1 - c_2(q)) \rho_2^+ e^{\rho_2^+ \epsilon} + (c_2(q) - 1) \rho_1^+ e^{\rho_1^+ \epsilon}} dy \\
&= \frac{\frac{q}{\delta^+} (\rho_2^+ - \rho_1^+) e^{(\mu_+ - \delta^+)y}}{(1 - c_2(q)) (\rho_2^+ - \rho_1^+)} dy \\
&= \frac{q}{\delta^+ (1 - c_2(q))} e^{(\mu_+ - \delta^+)y} dy,
\end{aligned}$$

where $\mathbb{P}(X_{e_q}^1 \in dy) = \frac{q}{\delta^+} e^{(\mu_+ - \delta^+)y} dy$. Substituting δ^+ and $c_2(q)$,

$$\int_0^\infty q e^{-qt} \mathbb{P}(X_t \in dy) dt = \frac{2(1+\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} e^{(\mu_+ - \sqrt{\mu_+^2 + 2q})y} dy. \quad (11)$$

When $y < \epsilon < 0$, for $\epsilon < 0$. Let $X_t^2 = \mu_- t + B_t$ be a Brownian motion and let τ_x^2 be the hitting time of process X_t^2 . For $y < \epsilon < 0$, by the Markov property and the memoryless property,

$$\begin{aligned} \mathbb{P}(X_{e_q} \in dy) &= \mathbb{P}(\tau_\epsilon < e_q, X_{e_q} \in dy) \\ &= \mathbb{P}(\tau_\epsilon < e_q) \mathbb{P}_\epsilon(X_{e_q} \in dy) \\ &= \mathbb{E} e^{-q\tau_\epsilon} \left(\mathbb{P}_\epsilon(\tau_0 < e_q, X_{e_q} \in dy) + \mathbb{P}_\epsilon(e_q < \tau_0, X_{e_q} \in dy) \right) \\ &= \mathbb{E} e^{-q\tau_\epsilon} \left(\mathbb{E}_\epsilon e^{-q\tau_0} \mathbb{P}(X_{e_q} \in dy) + \mathbb{P}_\epsilon(e_q < \tau_0^2, X_{e_q}^2 \in dy) \right). \end{aligned} \quad (12)$$

Solving the above equation we have

$$\begin{aligned}
\int_0^\infty qe^{-qt}\mathbb{P}(X_t \in dy)dt &= \mathbb{P}(X_{e_q} \in dy) \\
&= \lim_{\epsilon \rightarrow 0^-} \frac{\mathbb{E}e^{-q\tau_\epsilon}\mathbb{P}_\epsilon(e_q < \tau_0^2, X_{e_q}^2 \in dy)}{1 - \mathbb{E}e^{-q\tau_\epsilon}\mathbb{E}_\epsilon e^{-q\tau_0}} \\
&= \lim_{\epsilon \rightarrow 0^-} \frac{\mathbb{E}e^{-q\tau_\epsilon}\left(\mathbb{P}_\epsilon(X_{e_q}^2 \in dy) - \mathbb{P}_\epsilon(e_q > \tau_0^2)\mathbb{P}(X_{e_q}^2 \in dy)\right)}{1 - \mathbb{E}e^{-q\tau_\epsilon}\mathbb{E}_\epsilon e^{-q\tau_0}} \\
&= \lim_{\epsilon \rightarrow 0^-} \frac{\mathbb{E}e^{-q\tau_\epsilon}\left(\mathbb{P}_\epsilon(X_{e_q}^2 \in dy) - \mathbb{E}_\epsilon e^{-q\tau_0}\mathbb{P}(X_{e_q}^2 \in dy)\right)}{1 - \mathbb{E}e^{-q\tau_\epsilon}\mathbb{E}_\epsilon e^{-q\tau_0}} \\
&= \lim_{\epsilon \rightarrow 0^-} \frac{\mathbb{P}_\epsilon(X_{e_q}^2 \in dy) - e^{\rho_2^- \epsilon}\mathbb{P}(X_{e_q}^2 \in dy)}{(c_1(q) - 1)e^{\rho_2^- \epsilon} + (1 - c_1(q))e^{\rho_1^- \epsilon}} \quad l'H\hat{o}pital's\ rule \quad (13) \\
&= \lim_{\epsilon \rightarrow 0^-} \frac{\frac{d}{d\epsilon}\mathbb{P}_\epsilon(X_{e_q}^2 \in dy) - \rho_2^- e^{\rho_2^- \epsilon}\mathbb{P}(X_{e_q}^2 \in dy)}{(c_1(q) - 1)\rho_2^- e^{\rho_2^- \epsilon} + (1 - c_1(q))\rho_1^- e^{\rho_1^- \epsilon}} \\
&= \lim_{\epsilon \rightarrow 0^-} \frac{\frac{q}{\delta^-}(-\mu_- - \delta^-)e^{(\mu_- + \delta^-)(y - \epsilon)} - \rho_2^- e^{\rho_2^- \epsilon} \frac{q}{\delta^-} e^{(\mu_- + \delta^-)y}}{(c_1(q) - 1)\rho_2^- e^{\rho_2^- \epsilon} + (1 - c_1(q))\rho_1^- e^{\rho_1^- \epsilon}} dy \\
&= \frac{-\frac{q}{\delta^-}(\rho_2^- - \rho_1^-)e^{(\mu_- + \delta^-)y}}{-(1 - c_1(q))(\rho_2^- - \rho_1^-)} dy \\
&= \frac{q}{\delta^-(1 - c_1(q))} e^{(\mu_- + \delta^-)y} dy,
\end{aligned}$$

where $\mathbb{P}(X_{e_q}^2 \in dy) = \frac{q}{\delta^-} e^{(\mu_- + \delta^-)y} dy$. Substituting δ^- and $c_1(q)$,

$$\begin{aligned}
\int_0^\infty qe^{-qt}\mathbb{P}(X_t \in dy)dt &= \\
&= \frac{2(1 - \beta)q}{(1 + \beta)\mu_+ - (1 - \beta)\mu_- + (1 + \beta)\sqrt{\mu_+^2 + 2q} + (1 - \beta)\sqrt{\mu_-^2 + 2q}} e^{(\mu_- + \sqrt{\mu_-^2 + 2q})y} dy. \quad (14)
\end{aligned}$$

□

We can also identify the potential measure for SBM with arbitrary initial value.

Proposition 1.4. *The potential measure of the refracted skew Brownian motion X_t with skew level $a = 0$ and starting point x is given by:*

for $y > 0$

$$\mathbb{P}_x(X_{e_q} \in dy) = \begin{cases} \frac{qe^{\mu_+(y-x)}}{\sqrt{\mu_+^2+2q}} \left(e^{-|y-x|\sqrt{\mu_+^2+2q}} - e^{-(x+y)\sqrt{\mu_+^2+2q}} \right) dy \\ + \frac{2(1+\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2+2q} + (1-\beta)\sqrt{\mu_-^2+2q}} e^{\mu_+(y-x) - (x+y)\sqrt{\mu_+^2+2q}} dy, & x \geq 0, \\ \frac{2(1+\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2+2q} + (1-\beta)\sqrt{\mu_-^2+2q}} e^{(-\mu_- + \sqrt{\mu_-^2+2q})x} e^{(\mu_+ - \sqrt{\mu_+^2+2q})y} dy, & x \leq 0, \end{cases} \quad (15)$$

and for $y < 0$

$$\mathbb{P}_x(X_{e_q} \in dy) = \begin{cases} \frac{2(1-\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2+2q} + (1-\beta)\sqrt{\mu_-^2+2q}} e^{(-\mu_- + \sqrt{\mu_-^2+2q})x} e^{(\mu_- + \sqrt{\mu_-^2+2q})y} dy, & x \geq 0, \\ \frac{qe^{\mu_-(y-x)}}{\sqrt{\mu_-^2+2q}} \left(e^{-|y-x|\sqrt{\mu_-^2+2q}} - e^{(x+y)\sqrt{\mu_-^2+2q}} \right) dy \\ + \frac{2(1-\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2+2q} + (1-\beta)\sqrt{\mu_-^2+2q}} e^{\mu_-(y-x) + (x+y)\sqrt{\mu_-^2+2q}} dy, & x \leq 0. \end{cases} \quad (16)$$

Proof. CASE 1: $x \geq 0$ and $y > 0$,

$$\begin{aligned} \mathbb{P}_x(X_{e_q} \in dy) &= \mathbb{P}_x(X_{e_q} \in dy, e_q < \tau_0) + \mathbb{P}_x(X_{e_q} \in dy, e_q > \tau_0) \\ &= \mathbb{P}_x(X_{e_q}^1 \in dy, e_q < \tau_0^1) + \mathbb{P}_x(e_q > \tau_0) \mathbb{P}(X_{e_q} \in dy) \\ &= \mathbb{P}_x(X_{e_q}^1 \in dy) - \mathbb{P}_x(X_{e_q}^1 \in dy, e_q > \tau_0^1) + \mathbb{P}_x(e_q > \tau_0) \mathbb{P}(X_{e_q} \in dy) \\ &= \mathbb{P}_x(X_{e_q}^1 \in dy) - \mathbb{E}_x e^{-q\tau_0^1} \mathbb{P}(X_{e_q}^1 \in dy) + \mathbb{E}_x e^{-q\tau_0} \mathbb{P}(X_{e_q} \in dy) \\ &= \frac{qe^{\mu_+(y-x)}}{\sqrt{\mu_+^2+2q}} \left(e^{-|y-x|\sqrt{\mu_+^2+2q}} - e^{-(x+y)\sqrt{\mu_+^2+2q}} \right) dy \\ &\quad + \frac{2(1+\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2+2q} + (1-\beta)\sqrt{\mu_-^2+2q}} \\ &\quad \times e^{\mu_+(y-x) - (x+y)\sqrt{\mu_+^2+2q}} dy. \end{aligned}$$

CASE 2: $x \leq 0$ and $y > 0$,

$$\begin{aligned}
\mathbb{P}_x(X_{e_q} \in dy) &= \mathbb{P}_x(X_{e_q} \in dy, e_q > \tau_0) \\
&= \mathbb{P}_x(e_q > \tau_0) \mathbb{P}(X_{e_q} \in dy) \\
&= \mathbb{E}_x e^{-q\tau_0} \mathbb{P}(X_{e_q} \in dy) \\
&= \frac{2(1+\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} \\
&\quad \times e^{(-\mu_- + \sqrt{\mu_-^2 + 2q})x} e^{(\mu_+ - \sqrt{\mu_+^2 + 2q})y} dy.
\end{aligned}$$

CASE 3: $x \geq 0$ and $y < 0$,

$$\begin{aligned}
\mathbb{P}_x(X_{e_q} \in dy) &= \mathbb{P}_x(X_{e_q} \in dy, e_q > \tau_0) \\
&= \mathbb{P}_x(e_q > \tau_0) \mathbb{P}(X_{e_q} \in dy) \\
&= \mathbb{E}_x e^{-q\tau_0} \mathbb{P}(X_{e_q} \in dy) \\
&= \frac{2(1-\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} \\
&\quad \times e^{(-\mu_+ - \sqrt{\mu_+^2 + 2q})x} e^{(\mu_- + \sqrt{\mu_-^2 + 2q})y} dy.
\end{aligned}$$

CASE 4: $x \leq 0$ and $y > 0$

$$\begin{aligned}
\mathbb{P}_x(X_{e_q} \in dy) &= \mathbb{P}_x(X_{e_q} \in dy, e_q < \tau_0) + \mathbb{P}_x(X_{e_q} \in dy, e_q > \tau_0) \\
&= \mathbb{P}_x(X_{e_q}^2 \in dy, e_q < \tau_0^2) + \mathbb{P}_x(e_q > \tau_0) \mathbb{P}(X_{e_q} \in dy) \\
&= \mathbb{P}_x(X_{e_q}^2 \in dy) - \mathbb{P}_x(X_{e_q}^2 \in dy, e_q > \tau_0^2) + \mathbb{P}_x(e_q > \tau_0) \mathbb{P}(X_{e_q} \in dy) \\
&= \mathbb{P}_x(X_{e_q}^2 \in dy) - \mathbb{E}_x e^{-q\tau_0^2} \mathbb{P}(X_{e_q}^2 \in dy) + \mathbb{E}_x e^{-q\tau_0} \mathbb{P}(X_{e_q} \in dy) \\
&= \frac{qe^{\mu_-(y-x)}}{\sqrt{\mu_-^2 + 2q}} \left(e^{-|y-x|\sqrt{\mu_-^2 + 2q}} - e^{(x+y)\sqrt{\mu_-^2 + 2q}} \right) dy \\
&\quad + \frac{2(1-\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} \\
&\quad \times e^{\mu_-(y-x) + (x+y)\sqrt{\mu_-^2 + 2q}} dy.
\end{aligned}$$

□

Remark 2. Proposition 1.4 is obtained in (1.6) of [20]. However, note that there is a typo in the first term of (1.6) when $z > 0$ and $x \geq 0$: all terms in the exponential functions should involve μ .

We further consider the potential measure with barriers imposed. If $b_- < x < b_+$, then, starting from x with skew level is zero, the probability that X_{e_q} remains between b_- and b_+ is:

Proposition 1.5. Given $b_- < 0 < b_+$, $X_0 = x$, $a = 0$, the potential measure of the refracted skew Brownian motion with barriers at b_- and b_+ is given by

$$\mathbb{P}_x(X_{e_q} \in dy, e_q < \tau_{b_-} \wedge \tau_{b_+}) = \mathbb{P}_x(X_{e_q} \in dy) - \frac{w(x, b_+)}{w(b_-, b_+)} \mathbb{P}_{b_-}(X_{e_q} \in dy) - \frac{w(x, b_-)}{w(b_+, b_-)} \mathbb{P}_{b_+}(X_{e_q} \in dy)$$

and remaining below or above level b_- is:

$$\mathbb{P}_x(X_{e_q} \in dy, e_q < \tau_{b_-}) = \mathbb{P}_x(X_{e_q} \in dy) - \mathbb{E}_x e^{-q\tau_{b_-}} \mathbb{P}_{b_-}(X_{e_q} \in dy).$$

Proof.

$$\begin{aligned} & \mathbb{P}_x(X_{e_q} \in dy, e_q < \tau_{b_-} \wedge \tau_{b_+}) \\ &= \mathbb{P}_x(X_{e_q} \in dy) - \mathbb{P}_x(\tau_{b_-} < e_q \wedge \tau_{b_+}, X_{e_q} \in dy) - \mathbb{P}_x(\tau_{b_+} < e_q \wedge \tau_{b_-}, X_{e_q} \in dy) \\ &= \mathbb{P}_x(X_{e_q} \in dy) - \mathbb{P}_x(\tau_{b_-} < e_q \wedge \tau_{b_+}) \mathbb{P}_{b_-}(X_{e_q} \in dy) - \mathbb{P}_x(\tau_{b_+} < e_q \wedge \tau_{b_-}) \mathbb{P}_{b_+}(X_{e_q} \in dy) \\ &= \mathbb{P}_x(X_{e_q} \in dy) - \mathbb{E}_x(e^{-q\tau_{b_-}}, \tau_{b_-} < \tau_{b_+}) \mathbb{P}_{b_-}(X_{e_q} \in dy) - \mathbb{E}_x(e^{-q\tau_{b_+}}, \tau_{b_+} < \tau_{b_-}) \mathbb{P}_{b_+}(X_{e_q} \in dy) \\ &= \mathbb{P}_x(X_{e_q} \in dy) - \frac{w(x, b_+)}{w(b_-, b_+)} \mathbb{P}_{b_-}(X_{e_q} \in dy) - \frac{w(x, b_-)}{w(b_+, b_-)} \mathbb{P}_{b_+}(X_{e_q} \in dy). \end{aligned}$$

For the second part:

$$\begin{aligned} \mathbb{P}_x(X_{e_q} \in dy, e_q < \tau_{b_-}) &= \mathbb{P}_x(X_{e_q} \in dy) - \mathbb{P}_x(X_{e_q} \in dy, e_q > \tau_{b_-}) \\ &= \mathbb{P}_x(X_{e_q} \in dy) - \mathbb{P}_x(e_q > \tau_{b_-}) \mathbb{P}_{b_-}(X_{e_q} \in dy) \\ &= \mathbb{P}_x(X_{e_q} \in dy) - \mathbb{E}_x e^{-q\tau_{b_-}} \mathbb{P}_{b_-}(X_{e_q} \in dy). \end{aligned}$$

□

1.5 Transition density of an SBM with two-valued drift

In this section, we obtain the transition probability of an RSBM by inverting the Laplace transform.

We first define an auxiliary function

$$h(t; x, \mu) := \frac{|x|}{\sqrt{2\pi t^3}} e^{\frac{-(x+\mu t)^2}{2t}}, \quad x, \mu \in \mathbb{R}, t > 0, \quad (17)$$

where its Laplace transform for $x \neq 0$ is given by:

$$\int_0^\infty e^{-qt} h(t; x, \mu) dt = \exp \left[-(\mu + \operatorname{sgn}(x) \sqrt{2q + \mu^2})x \right], \quad (18)$$

and zero for $x = 0$. Additionally, we have the following properties for h :

$$h(t; x, \mu) = h(t; -x, -\mu), \quad h(t; x, -\mu) = h(t; -x, \mu) = e^{2\mu x} h(t; x, \mu),$$

and

$$h(t; x_1 + x_2, \mu) = \int_0^t h(t - \tau; x_1, \mu) h(\tau; x_2, \mu) d\tau,$$

see [72].

Theorem 1.6. *The density probability of the refracted skew Brownian motion, with skew level $a = 0$ and starting point $x = 0$ is given by:*

$$p(t; y) = \begin{cases} 2(1 + \beta) e^{2\mu_+ y} \int_0^\infty \int_0^t h(t - \tau; (1 + \beta)b + y, \mu_+) h(\tau; (1 - \beta)b, -\mu_-) d\tau db, & y > 0. \\ 2(1 - \beta) e^{2\mu_- y} \int_0^\infty \int_0^t h(t - \tau; (1 + \beta)b, \mu_+) h(\tau; (1 - \beta)b - y, -\mu_-) d\tau db, & y < 0. \end{cases} \quad (19)$$

Proof. For $y > 0$, taking the Laplace transform for $p(t; y)$ in (19) we have

$$\begin{aligned}
& \int_0^\infty e^{-qt} \mathbb{P}(X_t \in dy) dt = \int_0^\infty e^{-qt} p(t; y) dt dy \\
&= 2(1 + \beta) e^{2\mu_+ y} \int_0^\infty \int_0^\infty e^{-qt} \int_0^t h(t - \tau; (1 + \beta)b + y, \mu_+) h(\tau; (1 - \beta)b, -\mu_-) d\tau dt db dy \\
&= 2(1 + \beta) e^{2\mu_+ y} \int_0^\infty \left[\int_0^\infty e^{-qt} h(t; (1 + \beta)b + y, \mu_+) dt \right] \left[\int_0^\infty e^{-qt} h(t; (1 - \beta)b, -\mu_-) dt \right] db dy \\
&= 2(1 + \beta) e^{2\mu_+ y} \int_0^\infty \left[\exp \left(-(\mu_+ + \sqrt{2q + \mu_+^2})((1 + \beta)b + y) \right) \right. \\
&\quad \left. \left[\exp \left(-(-\mu_- + \sqrt{2q + \mu_-^2})((1 - \beta)b) \right) \right] db dy \right] \\
&= 2(1 + \beta) e^{2\mu_+ y} \int_0^\infty e^{\rho_1^+((1 + \beta)b + y)} e^{-\rho_2^- (1 - \beta)b} db dy \\
&= 2(1 + \beta) e^{2\mu_+ y} e^{\rho_1^+ y} \int_0^\infty e^{\rho_1^+ (1 + \beta)b} e^{-\rho_2^- (1 - \beta)b} db dy \\
&= 2(1 + \beta) e^{-\rho_2^+ y} \int_0^\infty e^{-(-(1 + \beta)\rho_1^+ + (1 - \beta)\rho_2^-)b} db dy \\
&= \frac{2(1 + \beta)}{-(1 + \beta)\rho_1^+ + (1 - \beta)\rho_2^-} e^{-\rho_2^+ y} dy \\
&= \frac{2(1 + \beta)}{(1 + \beta)[\mu_+ + \sqrt{2q + \mu_+^2}] + (1 - \beta)[- \mu_- + \sqrt{2q + \mu_-^2}]} e^{(\mu_+ - \sqrt{2q + \mu_+^2})y} dy,
\end{aligned}$$

which recovers (8).

Similarly, for $y < 0$,

$$\begin{aligned}
& \int_0^\infty e^{-qt} \mathbb{P}(X_t \in dy) dt = \int_0^\infty e^{-qt} p(t; y) dt dy \\
&= 2(1 - \beta) e^{2\mu_- y} \int_0^\infty \int_0^\infty e^{-qt} \int_0^t h(t - \tau; (1 + \beta)b, \mu_+) h(\tau; (1 - \beta)b - y, -\mu_-) d\tau dt db dy \\
&= 2(1 - \beta) e^{2\mu_- y} \int_0^\infty \left[\int_0^\infty e^{-qt} h(t; (1 + \beta)b, \mu_+) dt \right] \left[\int_0^\infty e^{-qt} h(t; (1 - \beta)b - y, -\mu_-) dt \right] db dy \\
&= 2(1 - \beta) e^{2\mu_- y} \int_0^\infty \left[\exp \left(-(\mu_+ + \sqrt{2q + \mu_+^2})(1 + \beta)b \right) \right. \\
&\quad \left. \left[\exp \left(-(-\mu_- + \sqrt{2q + \mu_-^2})((1 - \beta)b - y) \right) \right] \right] db dy \\
&= 2(1 - \beta) e^{2\mu_- y} \int_0^\infty e^{\rho_1^+(1+\beta)b} e^{-\rho_2^-((1-\beta)-y)b} db dy \\
&= 2(1 - \beta) e^{2\mu_- y} e^{\rho_2^- y} \int_0^\infty e^{\rho_1^+(1+\beta)b} e^{-\rho_2^-(1-\beta)b} db dy \\
&= 2(1 - \beta) e^{-\rho_1^- y} \int_0^\infty e^{-(-(1+\beta)\rho_1^+(1-\beta)\rho_2^-)b} db dy \\
&= \frac{2(1 - \beta)}{-(1 + \beta)\rho_1^+ + (1 - \beta)\rho_2^-} e^{-\rho_1^- y} dy \\
&= \frac{2(1 - \beta)}{(1 + \beta)[\mu_+ + \sqrt{2q + \mu_+^2}] + (1 - \beta)[- \mu_- + \sqrt{2q + \mu_-^2}]} e^{(\mu_- + \sqrt{2q + \mu_-^2})y} dy,
\end{aligned}$$

which recovers (8). □

Remark 3. *Our result aligns with the marginal density of x in the joint density provided in equation (2.4) in [48]. But it does not agree with equation (2.3) presented in [48], possibly due to a typo in (2.3) for the case $x > 0$. Although the same authors of [49] obtained the density function for the refracted skew Brownian motion containing alternating drifts using equation (2.3), their two results appear to be inconsistent, particularly in more general settings, see Section (1.10), for more details.*

The following corollaries concern special cases of Theorem 1.6, which appear in the previous work on SBM.

Corollary 1.1. *The transition probability of a skew Brownian motion with one constant drift,*

$\mu_- = \mu_+ = \mu$, with skew level $a = 0$ and starting point $x = 0$ is given by:

$$p(t; y) = \begin{cases} \frac{(1+\beta)}{\sqrt{2\pi t}} e^{-\frac{(y-\mu t)^2}{2t}} \left[1 - \frac{\beta\mu}{2} \sqrt{2\pi t} e^{\frac{(y+\beta\mu t)^2}{2t}} \operatorname{erfc}\left(\frac{y+\beta\mu t}{\sqrt{2t}}\right) \right] dy, & y > 0, \\ \frac{(1-\beta)}{\sqrt{2\pi t}} e^{-\frac{(y-\mu t)^2}{2t}} \left[1 - \frac{\beta\mu}{2} \sqrt{2\pi t} e^{\frac{(-y+\beta\mu t)^2}{2t}} \operatorname{erfc}\left(\frac{-y+\beta\mu t}{\sqrt{2t}}\right) \right] dy, & y < 0, \end{cases} \quad (20)$$

where $\operatorname{erfc}(z)$ is a complementary error function given by

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt.$$

The proof is deferred to the Section 1.8. Note that it recovers the previous result in [8]. For $\mu = 0$, it corresponds to a result in [110].

Corollary 1.2. *The transition probability of a skew Brownian motion with two alternating drifts,*

$-\mu_- = \mu_+ = \mu$, with skew level $a = 0$ and starting point $x = 0$ is given by:

$$p(t; y) = \begin{cases} \frac{(1+\beta)}{\sqrt{2\pi t}} e^{-\frac{(y-\mu t)^2}{2t}} \left[1 - \frac{\mu}{2} \sqrt{2\pi t} e^{\frac{(y+\mu t)^2}{2t}} \operatorname{erfc}\left(\frac{y+\mu t}{\sqrt{2t}}\right) \right] dy, & y > 0, \\ \frac{(1-\beta)}{\sqrt{2\pi t}} e^{-\frac{(y-\mu t)^2}{2t}} \left[1 - \frac{\mu}{2} \sqrt{2\pi t} e^{\frac{(-y+\mu t)^2}{2t}} \operatorname{erfc}\left(\frac{-y+\mu t}{\sqrt{2t}}\right) \right] dy, & y < 0. \end{cases} \quad (21)$$

Which corresponds to [49]. The proof is similar to the previous corollary so we omit it.

Corollary 1.3. *Define $f(t; y)$ as the following*

$$f(t; y) = \begin{cases} \int_0^\infty \int_0^t h(t - \tau; (1 + \beta)b + y, \mu_+) h(\tau; (1 - \beta)b, -\mu_-) d\tau db, & y > 0, \\ \int_0^\infty \int_0^t h(t - \tau; (1 + \beta)b, \mu_+) h(\tau; (1 - \beta)b - y, -\mu_-) d\tau db, & y < 0, \end{cases}$$

the Laplace transform of $f(t; y)$ is given by:

$$\hat{f}(q; y) = \begin{cases} \frac{1}{(1+\beta)[\mu_+ + \sqrt{2q+\mu_+^2}] + (1-\beta)[- \mu_- + \sqrt{2q+\mu_-^2}]} e^{(-\mu_+ - \sqrt{2q+\mu_+^2})y}, & y > 0, \\ \frac{1}{(1+\beta)[\mu_+ + \sqrt{2q+\mu_+^2}] + (1-\beta)[- \mu_- + \sqrt{2q+\mu_-^2}]} e^{(-\mu_- + \sqrt{2q+\mu_-^2})y}, & y < 0. \end{cases}$$

For simplicity we set $f^-(t; y)$ and $\hat{f}^-(q; y)$ to be the functions defined in Corollary 1.3 for $y < 0$, and set $f^+(t; y)$, $\hat{f}^+(q; y)$, when $y > 0$.

Proposition 1.7. *The cumulative distribution function (CDF) of the refracted skew Brownian motion with skew level $a = 0$ and initial $x = 0$ is given by: for $z < 0$:*

$$\begin{aligned} \mathbb{P}(Z \leq z) = & 2(1 - \beta) \int_0^\infty e^{2\mu_- (1-\beta)b} \left[\int_0^t \frac{1}{\sqrt{2\pi\tau}} h(t - \tau; (1 + \beta)b, \mu_+) \exp\left(\frac{-((1-\beta)b - z + \mu_- \tau)^2}{2\tau}\right) d\tau \right. \\ & \left. - \frac{\mu_-}{2} \int_0^t h(t - \tau; (1 + \beta)b, \mu_+) \operatorname{erfc}\left(\frac{(1-\beta)b - z + \mu_- \tau}{\sqrt{2\tau}}\right) d\tau \right] db, \end{aligned}$$

and for $z > 0$:

$$\begin{aligned} \mathbb{P}(Z \leq z) = & 1 - 2(1 + \beta) \int_0^\infty e^{-2\mu_+ (1+\beta)b} \left[\int_0^t \frac{1}{\sqrt{2\pi\tau}} h(t - \tau; (1 - \beta)b, -\mu_-) \exp\left(\frac{-((1+\beta)b + z - \mu_+ \tau)^2}{2\tau}\right) d\tau \right. \\ & \left. + \frac{\mu_+}{2} \int_0^t h(t - \tau; (1 - \beta)b, -\mu_-) \operatorname{erfc}\left(\frac{(1+\beta)b + z - \mu_+ \tau}{\sqrt{2\tau}}\right) d\tau \right] db. \end{aligned}$$

Based on Proposition 1.4 and Corollary 1.3 we can derive the transition probability of skew Brownian motion with two-valued drift when the starting point is any real value x .

Proposition 1.8. *The transition probability of the refracted skew Brownian motion, with skew level $a = 0$ and starting point x is given by:*

for $y > 0$

$$p(t; x, y) = \begin{cases} \left[\phi(y - x; \mu_+) - e^{-2\mu_+ x} \phi(x + y; \mu_+) \right] \\ + 2(1 + \beta) e^{2\mu_+ y} \int_0^\infty \int_0^t h(t - \tau; (1 + \beta)b + x + y, \mu_+) h(\tau; (1 - \beta)b, -\mu_-) d\tau db, & x \geq 0 \\ 2(1 + \beta) e^{2\mu_+ y} \int_0^\infty \int_0^t h(\tau; (1 + \beta)b + y, \mu_+) h(t - \tau; (1 - \beta)b - x, -\mu_-) d\tau db, & x < 0 \end{cases}$$

and for $y < 0$

$$p(t; x, y) = \begin{cases} 2(1 - \beta)e^{2\mu_- y} \int_0^\infty \int_0^t h(t - \tau; (1 + \beta)b + x, \mu_+) h(\tau; (1 - \beta)b - y, -\mu_-) d\tau db, & x > 0 \\ \left[\phi(y - x; \mu_-) - e^{2\mu_- y} \phi(-x - y; \mu_-) \right] \\ + 2(1 - \beta)e^{2\mu_- y} \int_0^\infty \int_0^t h(\tau; (1 + \beta)b, \mu_+) h(t - \tau; (1 - \beta)b - y - x, -\mu_-) d\tau db, & x \leq 0 \end{cases}$$

where $\phi(z; \mu)$ is density function of a Brownian motion with drift μ defined by:

$$\phi(z; \mu) = \frac{1}{\sqrt{2\pi t}} \exp \frac{-(z - \mu t)^2}{2t} dz.$$

The proof is provided in Section 1.9.

Corollary 1.4. *The transition probability of the refracted Brownian motion with starting point x is given by:*

For $y > 0$,

$$p(t; x, y) = \begin{cases} \left[\phi(y - x; \mu_+) - e^{-2\mu_+ x} \phi(x + y; \mu_+) \right] \\ + 2e^{2\mu_+ y} \int_0^\infty \int_0^t h(t - \tau; b + x + y, \mu_+) h(\tau; b, -\mu_-) d\tau db, & x \geq 0, \\ 2e^{2\mu_+ y} \int_0^\infty \int_0^t h(\tau; b + y, \mu_+) h(t - \tau; b - x, -\mu_-) d\tau db, & x < 0, \end{cases}$$

and for $y < 0$,

$$p(t; x, y) = \begin{cases} 2e^{2\mu_- y} \int_0^\infty \int_0^t h(t - \tau; b + x, \mu_+) h(\tau; b - y, -\mu_-) d\tau db, & x > 0, \\ \left[\phi(y - x; \mu_-) - e^{2\mu_- y} \phi(-x - y; \mu_-) \right] \\ + 2e^{2\mu_- y} \int_0^\infty \int_0^t h(\tau; b, \mu_+) h(t - \tau; b - y - x, -\mu_-) d\tau db, & x \leq 0. \end{cases}$$

This recovers the result in [71].

Proposition 1.9. *For the refracted skew Brownian motion with $a = 0$, the size of discontinuity for $p(t; 0, y)$ at $y = 0$ is given by:*

$$4\beta \int_0^\infty \int_0^t h(t - \tau; (1 + \beta)b, \mu_+) h(\tau; (1 - \beta)b, -\mu_-) d\tau db$$

If $\beta = 0$, i.e. there is no skewness in the process, the jump size is zero and the process is continuous. When $\mu_- = \mu_+ = \mu$, then the jump size is:

$$\frac{\sqrt{2}\beta}{\sqrt{\pi t}} e^{-\frac{\mu^2 t}{2}} \left(1 - \frac{\beta\mu}{2} \sqrt{2\pi t} \operatorname{erfc}\left(\frac{\beta\mu t}{\sqrt{2t}}\right) \right).$$

When $-\mu_- = \mu_+ = \mu$, i.e. alternating drifts, the jump size is:

$$\frac{\sqrt{2}\beta}{\sqrt{\pi t}} e^{-\frac{\mu^2 t}{2}} - \beta\mu \operatorname{erfc}\left(\frac{\mu t}{\sqrt{2t}}\right).$$

In both cases, when $\mu = 0$, the jump size is a linear function of β :

$$\frac{\sqrt{2}\beta}{\sqrt{\pi t}}, \quad t > 0.$$

It means that for fixed t , when $|\beta|$ is close to 1, the absolute value of the jump increases.

Based on alternating case we can argue that when $\mu_- \rightarrow -\infty$ and $\mu_+ \rightarrow \infty$, the jump size tends to 0, and when $\mu_- \rightarrow \infty$ and $\mu_+ \rightarrow -\infty$, the absolute value of jump size tends to ∞ .

1.6 Long term behaviors

In this section, we investigate the behavior of an RSBM as t approaches infinity. When $\mu_- > 0$ and $\mu_+ < 0$, as t grows larger and larger, the transition probability of the SBM converges to a smooth density that is independent of the starting point x .

Proposition 1.10. Given $\mu_- > 0$ and $\mu_+ < 0$, for any $x \in \mathbb{R}$ we have

$$\lim_{t \rightarrow \infty} p(t; x, y) = \begin{cases} \frac{-2(1+\beta)\mu_+\mu_-}{(1+\beta)\mu_- - (1-\beta)\mu_+} e^{2\mu_+y}, & y > 0, \\ \frac{-2(1-\beta)\mu_+\mu_-}{(1+\beta)\mu_- - (1-\beta)\mu_+} e^{2\mu_-y}, & y < 0. \end{cases}$$

Proof. We use the potential measure of SBM to prove this proposition since as q tends to 0, t tends to infinity. Suppose $\mu_- > 0$, $\mu_+ < 0$, from Proposition (1.4), when $y > 0$ and $x \geq 0$, we have

$$\begin{aligned} \lim_{t \rightarrow \infty} p(t; x, y) &= \lim_{q \rightarrow 0} \frac{1}{dy} \mathbb{P}_x(X_{e_q} \in dy) \\ &= \lim_{q \rightarrow 0} \frac{qe^{\mu_+(y-x)}}{\sqrt{\mu_+^2 + 2q}} \left(e^{-|y-x|\sqrt{\mu_+^2 + 2q}} - e^{-(x+y)\sqrt{\mu_+^2 + 2q}} \right) \\ &+ \lim_{q \rightarrow 0} \frac{2(1+\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} e^{\mu_+(y-x) - (x+y)\sqrt{\mu_+^2 + 2q}} \\ &= \lim_{q \rightarrow 0} \frac{2(1+\beta)e^{\mu_+(y-x) - (x+y)\sqrt{\mu_+^2 + 2q}} + 2(1+\beta)q \frac{d}{dq} \left(e^{\mu_+(y-x) - (x+y)\sqrt{\mu_+^2 + 2q}} \right)}{(1+\beta)(\mu_+^2 + 2q)^{-\frac{1}{2}} + (1-\beta)(\mu_-^2 + 2q)^{-\frac{1}{2}}} \quad \text{l'Hôpital's rule} \\ &= \frac{2(1+\beta)e^{\mu_+(y-x) - (x+y)|\mu_+|}}{(1+\beta)|\mu_+|^{-1} + (1-\beta)|\mu_-|^{-1}} \quad |\mu_+| = -\mu_+ \text{ and } |\mu_-| = \mu_- \\ &= \frac{-2(1+\beta)\mu_+\mu_-}{(1+\beta)\mu_- - (1-\beta)\mu_+} e^{2\mu_+y}. \end{aligned}$$

When $y > 0$ and $x \leq 0$, we have

$$\begin{aligned} \lim_{t \rightarrow \infty} p(t; x, y) &= \lim_{q \rightarrow 0} \frac{1}{dy} \mathbb{P}_x(X_{e_q} \in dy) \\ &= \lim_{q \rightarrow 0} \frac{2(1+\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} e^{(-\mu_- + \sqrt{\mu_-^2 + 2q})x} e^{(\mu_+ - \sqrt{\mu_+^2 + 2q})y} \\ &= \lim_{q \rightarrow 0} \frac{2(1+\beta)e^{(-\mu_- + \sqrt{\mu_-^2 + 2q})x} e^{(\mu_+ - \sqrt{\mu_+^2 + 2q})y} + 2(1+\beta)q \frac{d}{dq} \left(e^{(-\mu_- + \sqrt{\mu_-^2 + 2q})x} e^{(\mu_+ - \sqrt{\mu_+^2 + 2q})y} \right)}{(1+\beta)(\mu_+^2 + 2q)^{-\frac{1}{2}} + (1-\beta)(\mu_-^2 + 2q)^{-\frac{1}{2}}} \quad \text{l'Hôpital's rule} \\ &= \frac{2(1+\beta)e^{(-\mu_- + |\mu_-|)x} e^{(\mu_+ - |\mu_+|)y}}{(1+\beta)|\mu_+|^{-1} + (1-\beta)|\mu_-|^{-1}} \quad |\mu_+| = -\mu_+ \text{ and } |\mu_-| = \mu_- \\ &= \frac{-2(1+\beta)\mu_+\mu_-}{(1+\beta)\mu_- - (1-\beta)\mu_+} e^{2\mu_+y}. \end{aligned}$$

The desired result follows from a property of the Laplace transform.

When $y < 0$ and $x \geq 0$, we have

$$\begin{aligned}
\lim_{t \rightarrow \infty} p(t; x, y) &= \lim_{q \rightarrow 0} \frac{1}{dy} \mathbb{P}_x(X_{e_q} \in dy) \\
&= \lim_{q \rightarrow 0} \frac{2(1-\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} e^{(-\mu_+ - \sqrt{\mu_+^2 + 2q})x} e^{(\mu_- + \sqrt{\mu_-^2 + 2q})y} \\
&= \lim_{q \rightarrow 0} \frac{2(1-\beta)e^{(-\mu_+ - \sqrt{\mu_+^2 + 2q})x} e^{(\mu_- + \sqrt{\mu_-^2 + 2q})y} + 2(1-\beta)q \frac{d}{dq} \left(e^{(-\mu_+ - \sqrt{\mu_+^2 + 2q})x} e^{(\mu_- + \sqrt{\mu_-^2 + 2q})y} \right)}{(1+\beta)(\mu_+^2 + 2q)^{-\frac{1}{2}} + (1-\beta)(\mu_-^2 + 2q)^{-\frac{1}{2}}} \quad \text{l'Hôpital's rule} \\
&= \frac{2(1-\beta)e^{(-\mu_+ - |\mu_+|)x} e^{(\mu_- + |\mu_-|)y}}{(1+\beta)|\mu_+|^{-1} + (1-\beta)|\mu_-|^{-1}} \quad |\mu_+| = -\mu_+ \text{ and } |\mu_-| = \mu_- \\
&= \frac{-2(1-\beta)\mu_+\mu_-}{(1+\beta)\mu_- - (1-\beta)\mu_+} e^{2\mu_- y}.
\end{aligned}$$

When $y < 0$ and $x \leq 0$, we have

$$\begin{aligned}
\lim_{t \rightarrow \infty} p(t; x, y) &= \lim_{q \rightarrow 0} \frac{1}{dy} \mathbb{P}_x(X_{e_q} \in dy) \\
&= \lim_{q \rightarrow 0} \frac{qe^{\mu_-(y-x)}}{\sqrt{\mu_-^2 + 2q}} \left(e^{-|y-x|\sqrt{\mu_-^2 + 2q}} - e^{(x+y)\sqrt{\mu_-^2 + 2q}} \right) \\
&+ \lim_{q \rightarrow 0} \frac{2(1-\beta)q}{(1+\beta)\mu_+ - (1-\beta)\mu_- + (1+\beta)\sqrt{\mu_+^2 + 2q} + (1-\beta)\sqrt{\mu_-^2 + 2q}} e^{\mu_-(y-x) + (x+y)\sqrt{\mu_-^2 + 2q}} \\
&= \lim_{q \rightarrow 0} \frac{2(1-\beta)e^{\mu_-(y-x) + (x+y)\sqrt{\mu_-^2 + 2q}} + 2(1-\beta)q \frac{d}{dq} \left(e^{\mu_-(y-x) + (x+y)\sqrt{\mu_-^2 + 2q}} \right)}{(1+\beta)(\mu_+^2 + 2q)^{-\frac{1}{2}} + (1-\beta)(\mu_-^2 + 2q)^{-\frac{1}{2}}} \quad \text{l'Hôpital's rule} \\
&= \frac{2(1-\beta)e^{\mu_-(y-x) + (x+y)|\mu_-|}}{(1+\beta)|\mu_+|^{-1} + (1-\beta)|\mu_-|^{-1}} \quad |\mu_+| = -\mu_+ \text{ and } |\mu_-| = \mu_- \\
&= \frac{-2(1-\beta)\mu_+\mu_-}{(1+\beta)\mu_- - (1-\beta)\mu_+} e^{2\mu_- y}.
\end{aligned}$$

□

For $\mu_+ > 0$ and $\mu_- < 0$, the probability that an SBM, with sufficiently large t , tends towards either positive or negative infinity is determined by the following proposition.

Proposition 1.11. Given $\mu_+ > 0$ and $\mu_- < 0$, for any $x \in \mathbb{R}$ we have

$$\mathbb{P}_x(\lim_{t \rightarrow \infty} X_t = \infty) = \begin{cases} \frac{(1+\beta)\mu_+ e^{-2\mu_- x}}{(1+\beta)\mu_+ - (1-\beta)\mu_-}, & x < 0, \\ 1 + \frac{(1-\beta)\mu_- e^{-2\mu_+ x}}{(1+\beta)\mu_+ - (1-\beta)\mu_-}, & x > 0, \end{cases}$$

and

$$\mathbb{P}_x(\lim_{t \rightarrow \infty} X_t = -\infty) = \begin{cases} 1 - \frac{(1+\beta)\mu_+ e^{-2\mu_- x}}{(1+\beta)\mu_+ - (1-\beta)\mu_-}, & x < 0, \\ \frac{-(1-\beta)\mu_- e^{-2\mu_+ x}}{(1+\beta)\mu_+ - (1-\beta)\mu_-}, & x > 0. \end{cases}$$

Proof. For $x < 0 < b$, by Theorem 1.2 we have

$$\begin{aligned} \mathbb{P}_x \left(\sup_{0 < t < \infty} X_t \geq b \right) &= \mathbb{P}_x(\tau_b < \infty) = \lim_{q \rightarrow 0^+} \mathbb{E}_x[e^{-q\tau_b}] \\ &= \lim_{q \rightarrow 0} \frac{e^{\rho_2^- x}}{(1 - c_2(q))e^{\rho_2^+ b} + c_2(q)e^{\rho_1^+ b}} \\ &= \frac{(1 + \beta)\mu_+ e^{-2\mu_- x}}{(1 + \beta)\mu_+ - (1 - \beta)\mu_- + (1 - \beta)\mu_- e^{-2\mu_+ b}}. \end{aligned}$$

Then

$$\mathbb{P}_x \left(\sup_{0 < t < \infty} X_t = \infty \right) = \lim_{b \rightarrow \infty} \mathbb{P}_x \left(\sup_{0 < t < \infty} X_t \geq b \right) = \frac{(1 + \beta)\mu_+ e^{-2\mu_- x}}{(1 + \beta)\mu_+ - (1 - \beta)\mu_-}.$$

Similarly, for $0 < x < b$,

$$\begin{aligned} \mathbb{P}_x \left(\sup_{0 < t < \infty} X_t \geq b \right) &= \mathbb{P}_x(\tau_b < \infty) = \lim_{q \rightarrow 0^+} \mathbb{E}_x[e^{-q\tau_b}] \\ &= \lim_{q \rightarrow 0} \frac{(1 - c_2(q))e^{\rho_2^+ x} + c_2(q)e^{\rho_1^+ x}}{(1 - c_2(q))e^{\rho_2^+ b} + c_2(q)e^{\rho_1^+ b}} \\ &= \frac{(1 + \beta)\mu_+ - (1 - \beta)\mu_- + (1 - \beta)\mu_- e^{-2\mu_+ x}}{(1 + \beta)\mu_+ - (1 - \beta)\mu_- + (1 - \beta)\mu_- e^{-2\mu_+ b}}. \end{aligned}$$

Then

$$\mathbb{P}_x \left(\sup_{0 < t < \infty} X_t = \infty \right) = \lim_{b \rightarrow \infty} \mathbb{P}_x \left(\sup_{0 < t < \infty} X_t \geq b \right) = \frac{(1 + \beta)\mu_+ - (1 - \beta)\mu_- + (1 - \beta)\mu_- e^{-2\mu_+ x}}{(1 + \beta)\mu_+ - (1 - \beta)\mu_-}.$$

For $b \leq 0 \leq x$,

$$\begin{aligned} \mathbb{P}_x \left(\inf_{0 < t < \infty} X_t \leq b \right) &= \mathbb{P}_x(\tau_b < \infty) = \lim_{q \rightarrow 0^-} \mathbb{E}_x[e^{-q\tau_b}] \\ &= \lim_{q \rightarrow 0} \frac{e^{\rho_1^+ x}}{c_1(q)e^{\rho_2^- b} + (1 - c_1(q))e^{\rho_1^- b}} \\ &= \frac{-(1 - \beta)\mu_- e^{-2\mu_+ x}}{-(1 + \beta)\mu_+ e^{-2\mu_- b} + (1 + \beta)\mu_+ - (1 - \beta)\mu_-}. \end{aligned}$$

It follows that

$$\mathbb{P}_x \left(\inf_{0 < t < \infty} X_t = -\infty \right) = \lim_{b \rightarrow -\infty} \mathbb{P}_x \left(\inf_{0 < t < \infty} X_t \leq b \right) = \frac{-(1 - \beta)\mu_- e^{-2\mu_+ x}}{(1 + \beta)\mu_+ - (1 - \beta)\mu_-}.$$

For $b \leq x \leq 0$,

$$\begin{aligned} \mathbb{P}_x \left(\inf_{0 < t < \infty} X_t \leq b \right) &= \mathbb{P}_x(\tau_b < \infty) = \lim_{q \rightarrow 0^-} \mathbb{E}_x[e^{-q\tau_b}] \\ &= \lim_{q \rightarrow 0} \frac{c_1(q)e^{\rho_2^- x} + (1 - c_1(q))e^{\rho_1^- x}}{c_1(q)e^{\rho_2^- b} + (1 - c_1(q))e^{\rho_1^- b}} \\ &= \frac{-(1 + \beta)\mu_+ e^{-2\mu_- x} + (1 + \beta)\mu_+ - (1 - \beta)\mu_-}{-(1 + \beta)\mu_+ e^{-2\mu_- b} + (1 + \beta)\mu_+ - (1 - \beta)\mu_-}. \end{aligned}$$

So,

$$\mathbb{P}_x \left(\inf_{0 < t < \infty} X_t = -\infty \right) = \lim_{b \rightarrow -\infty} \mathbb{P}_x \left(\inf_{0 < t < \infty} X_t \leq b \right) = \frac{(1 + \beta)\mu_+ - (1 - \beta)\mu_- - (1 + \beta)\mu_+ e^{-2\mu_- x}}{(1 + \beta)\mu_+ - (1 - \beta)\mu_-}.$$

□

Proposition 1.12. For $\mu_- > 0$, $\mu_+ < 0$ and any real value x, z , we have $\mathbb{P}_x(\tau_z < \infty) = 1$.

Proof. We follow an argument similar to that in Proposition (1.11) but for $\mu_- > 0$ and $\mu_+ < 0$. For

$x < 0 < z$,

$$\mathbb{P}_x(\tau_z < \infty) = \lim_{q \rightarrow 0+} \mathbb{E}_x[e^{-q\tau_z}] = \lim_{q \rightarrow 0} \frac{e^{\rho_2^- x}}{(1 - c_2(q))e^{\rho_2^+ z} + c_2(q)e^{\rho_1^+ z}} = 1;$$

for $0 < x < z$,

$$\mathbb{P}_x(\tau_z < \infty) = \lim_{q \rightarrow 0+} \mathbb{E}_x[e^{-q\tau_z}] = \lim_{q \rightarrow 0} \frac{(1 - c_2(q))e^{\rho_2^+ x} + c_2(q)e^{\rho_1^+ x}}{(1 - c_2(q))e^{\rho_2^+ z} + c_2(q)e^{\rho_1^+ z}} = 1;$$

for $z \leq 0 \leq x$,

$$\mathbb{P}_x(\tau_z < \infty) = \lim_{q \rightarrow 0-} \mathbb{E}_x[e^{-q\tau_z}] = \lim_{q \rightarrow 0} \frac{e^{\rho_1^+ x}}{c_1(q)e^{\rho_2^- z} + (1 - c_1(q))e^{\rho_1^- z}} = 1;$$

and for $z \leq x \leq 0$,

$$\mathbb{P}_x(\tau_z < \infty) = \lim_{q \rightarrow 0-} \mathbb{E}_x[e^{-q\tau_z}] = \lim_{q \rightarrow 0} \frac{c_1(q)e^{\rho_2^- x} + (1 - c_1(q))e^{\rho_1^- x}}{c_1(q)e^{\rho_2^- z} + (1 - c_1(q))e^{\rho_1^- z}} = 1.$$

Note that for $\mu_- > 0$ and $\mu_+ < 0$, as $q \rightarrow 0$ both $c_1(q)$ and $c_2(q)$ tend to 1, and $\rho_1^-, \rho_2^-, \rho_1^+, \rho_2^+$ tends to $-2\mu_-, 0, 0, -2\mu_+$, respectively. $\square$

Proposition 1.13. *Given any real value x, z , $\mathbb{E}_x(\tau_z) < \infty$ for $\mu_- > 0$ and $\mu_+ < 0$, and it is infinite if either $\mu_+ = 0$ or $\mu_- = 0$.*

Proof. According to Theorem (1.2), if we define $\psi(q) = \mathbb{E}_x[e^{-q\tau_z}]$, then $\mathbb{E}_x[\tau_z] = -\psi'(0)$. Given $\mu_- > 0$ and $\mu_+ < 0$, when $x < 0 < z$,

$$\mathbb{E}_x[\tau_z] = \frac{2(1 + \beta)(\mu_- \mu_+ z - \mu_+^2 x) + (\mu_- (1 + \beta) - \mu_+ (1 - \beta)) (e^{-2\mu_+ z} - 1)}{2\mu_- \mu_+^2 (1 + \beta)};$$

when $0 < x < z$,

$$\mathbb{E}_x[\tau_z] = \frac{2(1 + \beta)\mu_- \mu_+ (z - x) + (\mu_- (1 + \beta) - \mu_+ (1 - \beta)) (e^{-2\mu_+ z} - e^{-2\mu_+ x})}{2\mu_- \mu_+^2 (1 + \beta)};$$

when $z \leq 0 \leq x$,

$$\mathbb{E}_x[\tau_z] = \frac{2(1-\beta)(\mu_- \mu_+ z - \mu_-^2 x) + (\mu_- (1+\beta) - \mu_+ (1-\beta)) (1 - e^{-2\mu_- z})}{2\mu_-^2 \mu_+ (1-\beta)}$$

and when $z \leq x \leq 0$,

$$\mathbb{E}_x[\tau_z] = \frac{2\mu_- \mu_+ (1-\beta)(z-x) + (\mu_- (1+\beta) - \mu_+ (1-\beta)) (e^{-2\mu_- x} - e^{-2\mu_- z})}{2\mu_-^2 \mu_+ (1-\beta)}.$$

The desired results then follow. □

1.7 Approximation and Simulation

As mentioned in the introduction, skew Brownian motion has many applications across various fields. A key aspect of these applications is the generation of samples from the distribution, which can be utilized in Monte Carlo simulations. However, due to the complexity of the stochastic differential equation (SDE) governing SBM and the intricate nature of its cumulative distribution function (CDF), traditional methods such as the Euler scheme or transformation methods are challenging to implement.

Although SBM does not have the independent increments because it is biased at the skew level, we can approximate its CDF on either side of the skew level; here is zero. In this section, we introduce two methods to generate quasi-random samples of skew Brownian motion. The first method approximates the CDF of SBM using a truncated normal distribution and then generates random samples based on the inverse of this approximated CDF. The second method employs a deep neural network to approximate the inverse of the CDF of SBM, subsequently generating random samples using the trained network.

1.7.1 Approximation by mixture of truncated normal distributions

A truncated normal distribution is a modification of the normal (Gaussian) distribution where values below a certain lower bound and/or above a certain upper bound are excluded, or "truncated." This results in a distribution that is confined to a specified range. The truncation can occur on either one side (left or right) or on both sides of the distribution. Suppose X is a normal random variable with mean μ and variance σ^2 , with probability density function $f(x)$ and cumulative distribution function $F(x)$. Then the PDF, $k(x)$, and CDF, $K(x)$, of the normal distribution truncated in $[a, b]$ is given by:

$$k(x) = \frac{f(x)}{F(b) - F(a)}, \quad K(x) = \frac{F(x) - F(a)}{F(b) - F(a)}.$$

Suppose $X = (X_t)_{t \geq 0}$ is the refracted skew Brownian motion, with skew level $a = 0$ and starting point $x = 0$ with the CDF given in Proposition (1.7). Given that the density function of X on either side of the skew level closely resembles a normal distribution, we assume that X is a truncated Brownian motion. So we approximate the CDF of RSBM with the following mixture of truncated normal distributions:

$$H(x) = \begin{cases} \alpha K_1(x), & x < 0, \\ \alpha + (1 - \alpha)K_2(x), & x > 0, \end{cases}$$

where $K_1(x) = \frac{F(x)}{F(0)}$ is the CDF of a truncated normal variable in support $(-\infty, 0)$, and $K_2(x) = \frac{G(x) - G(0)}{1 - G(0)}$ is the CDF of a truncated normal variable in support $(0, \infty)$. Here, F is a normal CDF with mean μ_1 , variance σ_1^2 and G is the normal CDF with mean μ_2 , variance σ_2^2 , and α is a constant in $[0, 1]$.

Given the parameters $\alpha, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2$ of H , we can generate random samples of the RSBM by transforming uniform random variables using the inverse CDF method. For a uniform random

variable u , the quasi-random sample RSBM is given by:

$$\hat{X} = \begin{cases} F^{-1}\left(\frac{u}{\alpha}F(0)\right), & u \leq \alpha, \\ G^{-1}\left(\frac{u-\alpha}{1-\alpha}(1-G(0)) + G(0)\right), & u > \alpha. \end{cases} \quad (22)$$

The remaining task is to determine the parameters of H . To achieve an accurate approximation, we use the Sequential Least Squares Programming (SLSQP) method [75] in the SciPy Python package. This optimization technique is employed to minimize the quantile of absolute errors between the RSBM CDF and H . Algorithm (1) outlines the necessary steps to generate random samples of RSBM using a mixture truncated normal distribution.

Algorithm 1 SBM Quasi-Random Sample Generation Using Truncated Normal Distribution

Require: $\mathbf{x}$: Vector of equispaced or non-equispaced values in the interval $[-A, A]$ with length N ,
params: Skew Brownian Motion parameters

Ensure: $\hat{X}$: Random sample of SBM

Calculate $\mathbb{P}(X \leq x_i)$ for each $i \in \{1, \dots, N\}$ based on Proposition (1.7),

Estimate parameters $\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha$ of the mixture truncated normal distribution by:

$$\min_i Q_{0.99}(|\mathbb{P}(X \leq x_i) - H(x_i)|),$$

Generate uniform random variables u ,

Generate SBM random samples $\hat{X}$ using the approximated mixture truncated normal distribution as per equation (22).

1.7.2 Using Deep Neural Network

A deep neural network (DNN) is a type of artificial neural network (ANN) with multiple layers between the input and output layers. These additional layers, often referred to as hidden layers, enable the network to model complex, hierarchical patterns in data. DNNs are a core component of deep learning, a subset of machine learning, and have many applications in computer vision, finance, games, etc. Researchers have shared an immense interest in applying neural networks to generate quasi-random samples from a distribution. For instance, Horger et al. in [63] introduced a neural network model capable of transforming samples from a uniform distribution into samples of any

specified probability density function. Hofert et al. in [62] presented a new approach for generating quasi-random samples from multivariate distributions with any underlying copula using generative neural networks. Furthermore, Gibson et al. in [53] proposed a method that employs Normalizing Flow generative models combined with Importance Sampling to simulate samples from rare event distributions and achieve accurate estimates. Here, we use DNN to approximate the inverse of CDF of an RSBM, then generate quasi-random samples using the trained network. Since deep neural networks (DNNs) are not the main focus of this paper, we will not delve into their details. Instead, we will briefly explain our algorithm and leave the rest to readers to explore further as needed.

Algorithm (2) explains the steps for training the neural network and generating random sample variables of RSBM. We use one node in both the input and output layers and employ four hidden layers with 512, 1025, 512, and 1024 nodes, respectively. Additionally, we utilize the Adam optimization algorithm. The number of epochs and iterations are 10,000 and 250 respectively.

Algorithm 2 SBM Quasi-Random Sample Generation Using DNN

Require: z : Equispaced or non-equispaced values in the interval $[-A, A]$,
params: Skew Brownian Motion parameters
Ensure: $\hat{Z}$: Random sample of RSBM
 Calculate $F(z) = \mathbb{P}(Z \leq z)$ based on Proposition (1.7).
for epochs **do**
 for Iterations **do**
 Get a chunk of $F(z)$
 Calculate the $\hat{z} = \hat{F}^{-1}(F(z))$ using the neural network
 Calculate the objective function (Mean Absolute Errors = $\frac{1}{N} \sum_i |z_i - \hat{z}_i|$)
 Update the neural network weights using back-propagation and auto-gradient
 end for
end for
 Generate uniform random variables u
 Generate SBM random samples $\hat{Z}$ using trained neural network and u as its input.

1.7.3 Simulation Results

In this section, we generate random samples of RSBM with different parameters using the two above-mentioned methods. In order to test the goodness-of-fit of generated random models, we employ a one-sample Kolmogorov–Smirnov (K-S) test. KS test is a nonparametric test to determine

whether a sample comes from a specified continuous distribution. The Kolmogorov–Smirnov statistic quantifies the distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution. Since the RSBM CDF is a right-continuous distribution with a single jump at the skew level, we can use a one-sample K-S test to determine whether the generated quasi-random samples follow the same distribution as the RSBM.

We consider four models for RSBM with different parameters (Table 1.1) and utilize the approximation methods to generate random samples. In addition, we consider $A = 20$ in both Algorithms (1 and 2).

Models	μ_-	μ_+	t	β
Model 1	-0.1	0.1	2	0.3
Model 2	2	-4	2	0.7
Model 3	1	3	1	-0.1
Model 4	-2	-3	3	0.9

Table 1.1: RSBM models with different parameters

Figures (1.1) and (1.2) present the random samples of RSBM generated using Deep Neural Network (DNN) and Truncated Normal Approximations (TNA), respectively. The first row of each figure compares the histogram of generated quasi-random samples with the theoretical PDF of RSBM, while the second row illustrates the empirical CDF of the generated samples against the theoretical CDF of the process. Note that the theoretical PDF and CDF are approximated using the Quadrature Adaptive General Segment (QAGS) algorithm as proposed in [98], which is available in the SciPy package in Python.

It can be observed that the generated samples are compatible with the theoretical distribution. Furthermore, the Kolmogorov-Smirnov (K-S) test indicates that the generated samples align well with the RSBM with a 95% confidence level. Table (1.2) displays the results of the K-S test for both approximation methods across different models, showing the effectiveness of each method in accurately generating samples that reflect the characteristics of the theoretical RSBM distribution.

Models	Method	K-S stats	p_value
Model 1	DNN	0.0143	0.2559
	TNA	0.0088	0.8305
Model 2	DNN	0.0116	0.5105
	TNA	0.0123	0.4277
Model 3	DNN	0.0106	0.6226
	TNA	0.0148	0.2181
Model 4	DNN	0.0096	0.7366
	TNA	0.0138	0.2874

Table 1.2: K-S Test Results for DNN and TNA Methods Across Different Models

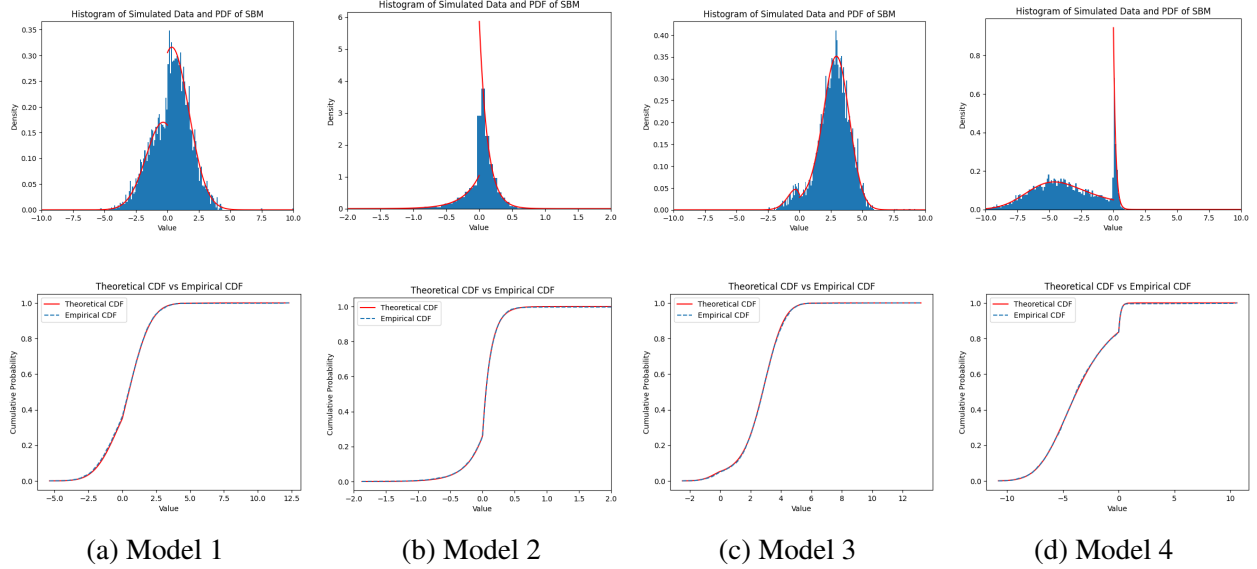


Figure 1.1: Histogram and CDF of generated random samples for different models with deep neural network against the exact PDF and CDF of the model

1.7.4 Application in Risk Measurement

The RSBM provides a versatile and powerful tool for modeling various asymmetric behaviors in financial markets. By incorporating skewness, it allows for more accurate pricing, risk management, and strategic decision-making, reflecting the true nature of financial assets and market dynamics. In markets, assets often have returns that are not symmetrically distributed. RSBM can better capture the empirical characteristics of asset returns, providing more realistic pricing and risk management models [11]. In addition, Investors with preferences for or against skewness in returns can benefit from portfolio optimization models incorporating skew Brownian motion. This allows for more

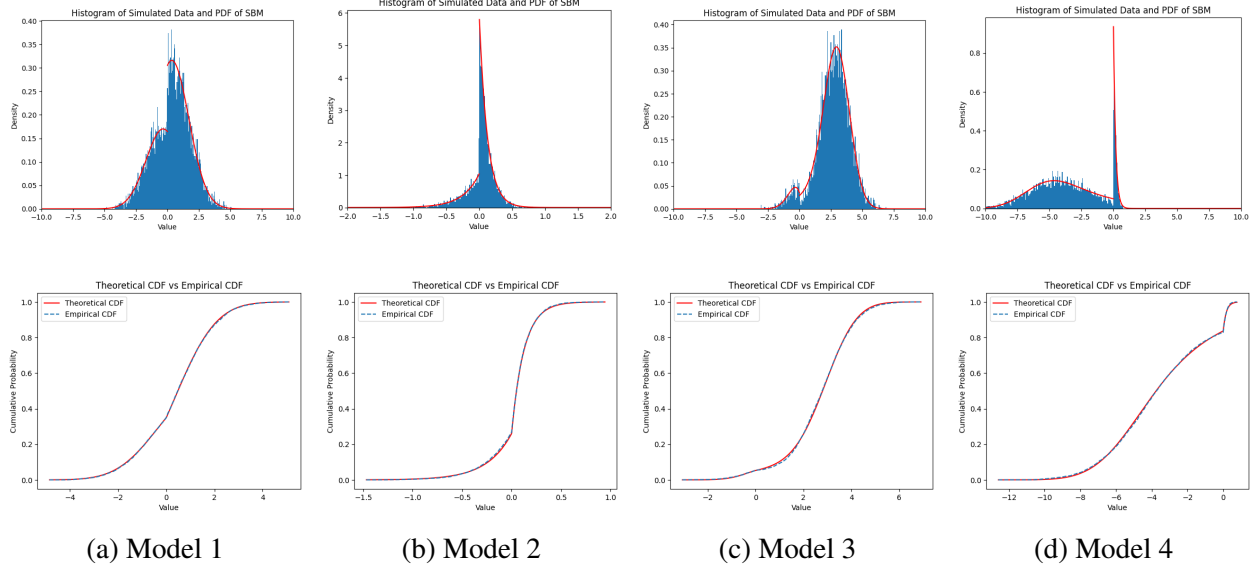


Figure 1.2: Histogram and CDF of generated random samples for different models with truncated Normal approximation against the exact PDF and CDF of the model

tailored portfolio construction, balancing risk and return according to the investor's skewness preferences. In the risk assessment aspect, conventional risk measures such as Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) can be adapted to account for skewness in the underlying return distributions, providing more accurate assessments of downside risk.

Calculating risk measures like VaR and CVaR based on the theoretical CDF of RSBM is challenging. Therefore, we use the truncated normal approximation (described in the previous section) to evaluate these risk measures based on the estimated parameters. Let L be the loss in a portfolio that follows an RSBM, for a confidence level q , $\text{VaR}_q(L)$ defined as:

$$\mathbb{P}(L \geq \text{VaR}_q(L)) = 1 - q.$$

One can easily calculate the VaR based on estimated parameters of mixed truncated normal approximation with:

$$\text{VaR}_q(L) = \begin{cases} \mu_1 + \sigma_1 \Phi^{-1} \left(\frac{q}{\alpha} \Phi \left(\frac{-\mu_1}{\sigma_1} \right) \right), & q \leq \alpha, \\ \mu_2 + \sigma_2 \Phi^{-1} \left(\frac{q-\alpha}{1-\alpha} + \frac{1-q}{1-\alpha} \Phi \left(\frac{-\mu_2}{\sigma_2} \right) \right), & q > \alpha, \end{cases} \quad (23)$$

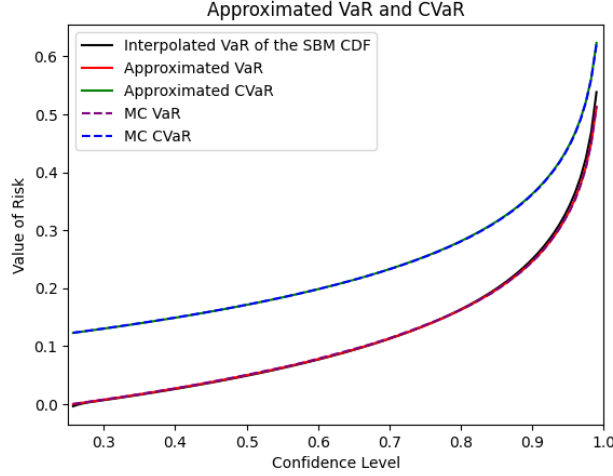


Figure 1.3: Evaluating approximated VaR and CVaR using mixed truncated normal distribution and Monte-Carlo (MC) simulation methods for values greater than $\alpha = 0.2569$.

Since the VaR of loss values is a positive value, the CVaR can be calculated as follows:

$$\begin{aligned}
 \text{CVaR}_q(L) &= \mathbb{E}[L|L \geq \text{VaR}_q(L)] = \frac{1}{1-q} \int_{\text{VaR}_q(L)}^{\infty} xh(x)dx \\
 &= \frac{1}{1-q} \int_{\text{VaR}_q(L)}^{\infty} \frac{1-\alpha}{1-G(0)} xg(x)dx \\
 &= \frac{1-\alpha}{(1-\Phi(\frac{-\mu_2}{\sigma_2}))(1-q)} \int_{\text{VaR}_q(L)}^{\infty} xg(x)dx \\
 &= \mu_2 + \frac{\sigma_2(1-\alpha)}{(1-\Phi(\frac{-\mu_2}{\sigma_2}))(1-q)} \phi \left(\Phi^{-1} \left(\frac{q-\alpha}{1-\alpha} + \frac{1-q}{1-\alpha} \Phi \left(\frac{-\mu_2}{\sigma_2} \right) \right) \right).
 \end{aligned} \tag{24}$$

We utilize Model 2 in Table (1.1) to evaluate the accuracy of the approximated VaR and CVaR. First, we compute the CDF of the RSBM using 2500 data points from the interval $[-20, 20]$. We next calculate the VaR using the linear interpolation method on the CDF and data points. We then use the estimated parameters of the mixed truncated distribution to approximate the VaR and CVaR. Finally, we run 100,000 quasi-random samples through the historical Monte Carlo simulation to calculate the VaR and CVaR for different confidence levels greater than $\alpha = 0.2569$ which acts like a border to have positive VaR. Note that we do not report negative VaR and CVaR because negative loss is not considered a risk.

1.8 Proof of Corollary 1.1

For $\mu_- = \mu_+ = \mu$ and $y > 0$,

$$\begin{aligned}
& \mathbb{P}(X_t \in dy) \\
&= 2(1+\beta)e^{2\mu y} \int_0^\infty \int_0^t h(t-\tau; (1+\beta)b+y, \mu) h(\tau; (1-\beta)b, -\mu) d\tau db dy \\
&= 2(1+\beta)e^{2\mu y} \int_0^\infty \int_0^t h(t-\tau; (1+\beta)b+y, \mu) h(\tau; (1-\beta)b, \mu) e^{2\mu(1-\beta)b} d\tau db dy \\
&= 2(1+\beta)e^{2\mu y} \int_0^\infty e^{2\mu(1-\beta)b} h(t-\tau; (1+\beta)b+y+(1-\beta)b, \mu) db dy \\
&= 2(1+\beta)e^{2\mu y} \int_0^\infty e^{2\mu(1-\beta)b} h(t; y+2b, \mu) db dy \\
&= \frac{2(1+\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} \int_0^\infty e^{2\mu(1-\beta)b} (y+2b) e^{\frac{-(y+2b+\mu t)^2}{2t}} db dy \\
&= \frac{2(1+\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} \int_0^\infty (y+2b) e^{\frac{-y^2-(2b)^2-\mu^2 t^2-2(2b)y-2y\mu t-2(2b)\mu t+4(1-\beta)b\mu t}{2t}} db dy \\
&= \frac{2(1+\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} e^{\frac{-(y+\mu t)^2}{2t}} \int_0^\infty (y+2b) e^{\frac{-(2b)^2-2(2b)y-2(2b)\beta\mu t-(y+\beta\mu t)^2+(y+\beta\mu t)^2}{2t}} db dy \\
&= \frac{2(1+\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} e^{\frac{-(y+\mu t)^2}{2t}} \int_0^\infty (y+2b) e^{\frac{-(2b+y+\beta\mu t)^2}{2t}} e^{\frac{(y+\beta\mu t)^2}{2t}} db dy \\
&= \frac{2(1+\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} e^{\frac{-(y+\mu t)^2}{2t}} e^{\frac{(y+\beta\mu t)^2}{2t}} \int_0^\infty (y+2b) e^{\frac{-(2b+y+\beta\mu t)^2}{2t}} db dy \\
&= \frac{2(1+\beta)}{\sqrt{2\pi t^3}} e^{\frac{-(y-\mu t)^2}{2t}} e^{\frac{(y+\beta\mu t)^2}{2t}} \left[\int_0^\infty (y+2b+\beta\mu t) e^{\frac{-(2b+y+\beta\mu t)^2}{2t}} db - \beta\mu t \int_0^\infty e^{\frac{-(2b+y+\beta\mu t)^2}{2t}} db \right] dy \\
&= \frac{(1+\beta)}{\sqrt{2\pi t^3}} e^{\frac{-(y-\mu t)^2}{2t}} e^{\frac{(y+\beta\mu t)^2}{2t}} \left[\underbrace{\int_y^\infty (v+\beta\mu t) e^{\frac{-(v+\beta\mu t)^2}{2t}} dv}_{I_1} - \beta\mu t \underbrace{\int_y^\infty e^{\frac{-(v+\beta\mu t)^2}{2t}} dv}_{I_2} \right] dy \\
&= \frac{(1+\beta)}{\sqrt{2\pi t^3}} e^{\frac{-(y-\mu t)^2}{2t}} e^{\frac{(y+\beta\mu t)^2}{2t}} \left[te^{\frac{-(y+\beta\mu t)^2}{2t}} - \beta\mu t \sqrt{2t} \frac{\sqrt{\pi}}{2} \operatorname{erfc}\left(\frac{y+\beta\mu t}{\sqrt{2t}}\right) \right] dy \\
&= \frac{(1+\beta)}{\sqrt{2\pi t}} e^{\frac{-(y-\mu t)^2}{2t}} e^{\frac{(y+\beta\mu t)^2}{2t}} \left[e^{\frac{-(y+\beta\mu t)^2}{2t}} - \frac{\beta\mu}{2} \sqrt{2\pi t} \operatorname{erfc}\left(\frac{y+\beta\mu t}{\sqrt{2t}}\right) \right] dy \\
&= \frac{(1+\beta)}{\sqrt{2\pi t}} e^{\frac{-(y-\mu t)^2}{2t}} \left[1 - \frac{\beta\mu}{2} \sqrt{2\pi t} e^{\frac{(y+\beta\mu t)^2}{2t}} \operatorname{erfc}\left(\frac{y+\beta\mu t}{\sqrt{2t}}\right) \right] dy,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \int_y^\infty (v + \beta\mu t) e^{\frac{-(v+\beta\mu t)^2}{2t}} dv \\
&= t \int_{\left(\frac{y+\beta\mu t}{\sqrt{2t}}\right)^2}^\infty e^{-u} du \quad u = \left(\frac{v + \beta\mu t}{\sqrt{2t}}\right)^2, \quad t du = (v + \beta\mu t) dv \\
&= t e^{\frac{-(y+\beta\mu t)^2}{2t}},
\end{aligned}$$

and

$$\begin{aligned}
I_2 &= \int_y^\infty e^{\frac{-(v+\beta\mu t)^2}{2t}} dv \\
&= \sqrt{2t} \int_{\frac{y+\beta\mu t}{\sqrt{2t}}}^\infty e^{-u^2} du \\
&= \sqrt{2t} \frac{\sqrt{\pi}}{2} \operatorname{erfc}\left(\frac{y + \beta\mu t}{\sqrt{2t}}\right).
\end{aligned}$$

For $\mu_- = \mu_+ = \mu$ and $y < 0$,

$$\begin{aligned}
& \mathbb{P}(X_t \in dy) \\
&= 2(1-\beta)e^{2\mu y} \int_0^\infty \int_0^t h(t-\tau; (1+\beta)b, \mu) h(\tau; (1-\beta)b-y, -\mu) d\tau db dy \\
&= 2(1-\beta)e^{2\mu y} \int_0^\infty \int_0^t h(t-\tau; (1+\beta)b, -\mu) h(\tau; (1-\beta)b-y, -\mu) e^{-2\mu(1+\beta)b} d\tau db dy \\
&= 2(1-\beta)e^{2\mu y} \int_0^\infty e^{-2\mu(1+\beta)b} h(t-\tau; (1+\beta)b + (1-\beta)b-y, -\mu) db dy \\
&= 2(1-\beta)e^{2\mu y} \int_0^\infty e^{-2\mu(1+\beta)b} h(t; -y+2b, -\mu) db dy \\
&= \frac{2(1-\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} \int_0^\infty e^{-2\mu(1+\beta)b} (-y+2b) e^{\frac{-(-y+2b-\mu t)^2}{2t}} db dy \\
&= \frac{2(1-\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} \int_0^\infty (-y+2b) e^{\frac{-y^2-(2b)^2-\mu^2 t^2+2(2b)y-2y\mu t+2(2b)\mu t-4(1+\beta)b\mu t}{2t}} db dy \\
&= \frac{2(1-\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} e^{\frac{-(y+\mu t)^2}{2t}} \int_0^\infty (-y+2b) e^{\frac{-(2b)^2+2(2b)y-2(2b)\beta\mu t-(-y+\beta\mu t)^2+(-y+\beta\mu t)^2}{2t}} db dy \\
&= \frac{2(1-\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} e^{\frac{-(y+\mu t)^2}{2t}} \int_0^\infty (-y+2b) e^{\frac{-(2b-y+\beta\mu t)^2}{2t}} e^{\frac{(-y+\beta\mu t)^2}{2t}} db dy \\
&= \frac{2(1-\beta)}{\sqrt{2\pi t^3}} e^{2\mu y} e^{\frac{-(y+\mu t)^2}{2t}} e^{\frac{(-y+\beta\mu t)^2}{2t}} \int_0^\infty (-y+2b) e^{\frac{-(-y+2b+\beta\mu t)^2}{2t}} db dy \\
&= \frac{2(1-\beta)}{\sqrt{2\pi t^3}} e^{\frac{-(y-\mu t)^2}{2t}} e^{\frac{(-y+\beta\mu t)^2}{2t}} \left[\int_0^\infty (-y+2b+\beta\mu t) e^{\frac{-(-y+2b+\beta\mu t)^2}{2t}} db - \beta\mu t \int_0^\infty e^{\frac{-(-y+2b+\beta\mu t)^2}{2t}} db \right] dy \\
&= \frac{(1-\beta)}{\sqrt{2\pi t^3}} e^{\frac{-(y-\mu t)^2}{2t}} e^{\frac{(-y+\beta\mu t)^2}{2t}} \left[\underbrace{\int_{-y}^\infty (v+\beta\mu t) e^{\frac{-(v+\beta\mu t)^2}{2t}} dv}_{I_1} - \beta\mu t \underbrace{\int_{-y}^\infty e^{\frac{-(v+\beta\mu t)^2}{2t}} dv}_{I_2} \right] dy \\
&= \frac{(1-\beta)}{\sqrt{2\pi t^3}} e^{\frac{-(y-\mu t)^2}{2t}} e^{\frac{(-y+\beta\mu t)^2}{2t}} \left[te^{\frac{-(-y+\beta\mu t)^2}{2t}} - \beta\mu t \sqrt{2t} \frac{\sqrt{\pi}}{2} \operatorname{erfc}\left(\frac{-y+\beta\mu t}{\sqrt{2t}}\right) \right] dy \\
&= \frac{(1-\beta)}{\sqrt{2\pi t}} e^{\frac{-(y-\mu t)^2}{2t}} e^{\frac{(-y+\beta\mu t)^2}{2t}} \left[e^{\frac{-(-y+\beta\mu t)^2}{2t}} - \frac{\beta\mu}{2} \sqrt{2\pi t} \operatorname{erfc}\left(\frac{-y+\beta\mu t}{\sqrt{2t}}\right) \right] dy \\
&= \frac{(1-\beta)}{\sqrt{2\pi t}} e^{\frac{-(y-\mu t)^2}{2t}} \left[1 - \frac{\beta\mu}{2} \sqrt{2\pi t} e^{\frac{(-y+\beta\mu t)^2}{2t}} \operatorname{erfc}\left(\frac{-y+\beta\mu t}{\sqrt{2t}}\right) \right] dy,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \int_{-y}^{\infty} (v + \beta\mu t) e^{\frac{-(v+\beta\mu t)^2}{2t}} dv \\
&= t \int_{(\frac{-y+\beta\mu t}{\sqrt{2t}})^2}^{\infty} e^{-u} du \quad u = \left(\frac{v + \beta\mu t}{\sqrt{2t}}\right)^2, \quad t du = (v + \beta\mu t) dv \\
&= t e^{\frac{-(-y+\beta\mu t)^2}{2t}},
\end{aligned}$$

and

$$\begin{aligned}
I_2 &= \int_{-y}^{\infty} e^{\frac{-(v+\beta\mu t)^2}{2t}} dv \\
&= \sqrt{2t} \int_{\frac{-y+\beta\mu t}{\sqrt{2t}}}^{\infty} e^{-u^2} du \\
&= \sqrt{2t} \frac{\sqrt{\pi}}{2} \operatorname{erfc}\left(\frac{-y + \beta\mu t}{\sqrt{2t}}\right).
\end{aligned}$$

1.9 Proof of Proposition 1.8

In this part of the appendix, we find the general transition probability in Proposition (1.8) by inverting the Laplace transform.

Remark 4. By formula 183 in table 2.3 in [3], and a property of Laplace transform one can show that

$$\int_0^{\infty} e^{-st} \frac{1}{\sqrt{2\pi t}} e^{\frac{-(k-\mu t)^2}{2t} - \mu k} dt = \frac{1}{\sqrt{2s + \mu^2}} e^{-|k| \sqrt{2s + \mu^2}}, \quad (25)$$

which is equivalent to formulas 1.0.5 and 1.0.6 in [21].

In addition, based on formula 181 in table 2.3 in [3], and using a property of Laplace transform one can show that for $k \neq 0$,

$$\int_0^{\infty} e^{-st} h(t; k, \mu) e^{\mu k} dt = e^{-|k| \sqrt{2s + \mu^2}}, \quad (26)$$

which is equivalent to formulas 2.0.1 and 2.0.2 in [21] and equation (18).

Now using the above remark and Corollary 1.3, we can calculate the inverse Laplace transform of $\mathbb{P}_x(X_{e_q} \in dy)$ in Proposition (1.4).

Proof. For $y > 0$ and $x \geq 0$,

$$\begin{aligned} p(x, y, t) &= \mathcal{L}^{-1}\{\mathbb{P}_x(X_{e_q} \in dy)\} \\ &= e^{\mu_+(y-x)} \left[\phi(y-x; \mu_+) e^{-\mu_+(y-x)} - \phi(x+y; \mu_+) e^{-\mu_+(x+y)} \right] + 2(1+\beta) e^{2\mu_+y} f^+(t; y) * h(t; x, \mu_+) \\ &= \phi(y-x; \mu_+) - e^{-2\mu_+x} \phi(x+y; \mu_+) + 2(1+\beta) e^{2\mu_+y} f^+(t; y) * h(t; x, \mu_+) \end{aligned}$$

For the last term:

$$\begin{aligned} & 2(1+\beta) e^{2\mu_+y} \int_0^t f^+(\eta; y) h(t-\eta; x, \mu_+) d\eta \\ &= 2(1+\beta) e^{2\mu_+y} \int_0^\infty \int_0^t \int_0^\eta h(\eta-\tau; (1+\beta)b+y, \mu_+) h(\tau; (1-\beta)b, -\mu_-) h(t-\eta; x, \mu_+) d\tau d\eta db \\ &= 2(1+\beta) e^{2\mu_+y} \int_0^\infty \int_0^t h(\tau; (1-\beta)b, -\mu_-) \int_\tau^t h(\eta-\tau; (1+\beta)b+y, \mu_+) h(t-\eta; x, \mu_+) d\eta d\tau db \\ &= 2(1+\beta) e^{2\mu_+y} \int_0^\infty \int_0^t h(\tau; (1-\beta)b, -\mu_-) \int_0^{t-\tau} h(t-\tau-u; (1+\beta)b+y, \mu_+) h(u; x, \mu_+) du d\tau db \\ &= 2(1+\beta) e^{2\mu_+y} \int_0^\infty \int_0^t h(t-\tau; (1+\beta)b+x+y, \mu_+) h(\tau; (1-\beta)b, -\mu_-) d\tau db. \end{aligned}$$

We change the order of τ and η in the second equality and use changing of variables $u = t - \eta$, in the third equality.

So,

$$\begin{aligned} p(x, y, t) &= \phi(y-x; \mu_+) - e^{-2\mu_+x} \phi(x+y; \mu_+) \\ &\quad + 2(1+\beta) e^{2\mu_+y} \int_0^\infty \int_0^t h(t-\tau; (1+\beta)b+x+y, \mu_+) h(\tau; (1-\beta)b, -\mu_-) d\tau db. \end{aligned}$$

For $y > 0$ and $x \leq 0$,

$$\begin{aligned}
p(x, y, t) &= \mathcal{L}^{-1}\{\mathbb{P}_x(X_{e_q} \in dy)\} \\
&= 2(1 + \beta)e^{2\mu+y} f^+(t; y) * h(t; -x, \mu_-) e^{-2\mu-x} \\
&= 2(1 + \beta)e^{2\mu+y} \int_0^t f^+(\eta; y) h(t - \eta; -x, -\mu_-) d\eta \\
&= 2(1 + \beta)e^{2\mu+y} \int_0^\infty \int_0^t \int_0^\eta h(\tau; (1 + \beta)b + y, \mu_+) h(\eta - \tau; (1 - \beta)b, -\mu_-) h(t - \eta; -x, -\mu_-) d\tau d\eta db \\
&= 2(1 + \beta)e^{2\mu+y} \int_0^\infty \int_0^t h(\tau; (1 + \beta)b + y, \mu_+) \int_\tau^t h(\eta - \tau; (1 - \beta)b, -\mu_-) h(t - \eta; -x, -\mu_-) d\eta d\tau db \\
&= 2(1 + \beta)e^{2\mu+y} \int_0^\infty \int_0^t h(\tau; (1 + \beta)b + y, \mu_+) \int_0^{t-\tau} h(t - \tau - u; (1 - \beta)b, -\mu_-) h(u; -x, -\mu_-) du d\tau db \\
&= 2(1 + \beta)e^{2\mu+y} \int_0^\infty \int_0^t h(\tau; (1 + \beta)b + y, \mu_+) h(t - \tau; (1 - \beta)b - x, -\mu_-) d\tau db.
\end{aligned}$$

We change the order of τ and η in the fifth equality and use the change of variable $u = t - \eta$ in the sixth equality.

For $y < 0$ and $x \geq 0$,

$$\begin{aligned}
p(x, y, t) &= \mathcal{L}^{-1}\{\mathbb{P}_x(X_{e_q} \in dy)\} \\
&= 2(1 - \beta)e^{2\mu-y} f^-(t; y) * h(t; x, \mu_+) \\
&= 2(1 - \beta)e^{2\mu-y} \int_0^t f^-(\eta; y) h(t - \eta; x, \mu_+) d\eta \\
&= 2(1 - \beta)e^{2\mu-y} \int_0^\infty \int_0^t \int_0^\eta h(\eta - \tau; (1 + \beta)b, \mu_+) h(\tau; (1 - \beta)b - y, -\mu_-) h(t - \eta; x, \mu_+) d\tau d\eta db \\
&= 2(1 - \beta)e^{2\mu-y} \int_0^\infty \int_0^t h(\tau; (1 - \beta)b - y, -\mu_-) \int_\tau^t h(\eta - \tau; (1 + \beta)b, \mu_+) h(t - \eta; x, \mu_+) d\eta d\tau db \\
&= 2(1 - \beta)e^{2\mu-y} \int_0^\infty \int_0^t h(\tau; (1 - \beta)b - y, -\mu_-) \int_0^{t-\tau} h(t - \tau - u; (1 + \beta)b, \mu_+) h(u; x, \mu_+) du d\tau db \\
&= 2(1 - \beta)e^{2\mu-y} \int_0^\infty \int_0^t h(t - \tau; (1 + \beta)b + x, \mu_+) h(\tau; (1 - \beta)b - y, -\mu_-) d\tau db.
\end{aligned}$$

We change the order of integrals in the fifth equality and use changing the variable $u = t - \eta$ in the sixth equality.

For $y < 0$ and $x \leq 0$,

$$\begin{aligned}
p(x, y, t) &= \mathcal{L}^{-1}\{\mathbb{P}_x(X_{e_q} \in dy)\} \\
&= e^{\mu_-(y-x)} \left[\phi(y-x; \mu_-) e^{-\mu_-(y-x)} - \phi(-x-y; \mu_-) e^{\mu_-(x+y)} \right] \\
&\quad + 2(1-\beta) e^{2\mu_-y} f^-(t; y) * h(t; -x, \mu_-) e^{-2\mu_-x} \\
&= \phi(y-x; \mu_-) - e^{2\mu_-y} \phi(-x-y; \mu_-) + 2(1-\beta) e^{2\mu_-y} f^-(t; y) * h(t; -x, -\mu_-).
\end{aligned}$$

For the last term:

$$\begin{aligned}
& 2(1-\beta) e^{2\mu_-y} \int_0^t f^-(\eta; y) h(t-\eta; -x, -\mu_-) d\eta \\
&= 2(1-\beta) e^{2\mu_-y} \int_0^\infty \int_0^t \int_0^\eta h(\tau; (1+\beta)b, \mu_+) h(\eta-\tau; (1-\beta)b-y, -\mu_-) h(t-\eta; -x, -\mu_-) d\tau d\eta db \\
&= 2(1-\beta) e^{2\mu_-y} \int_0^\infty \int_0^t h(\tau; (1+\beta)b, \mu_+) \int_\tau^t h(\eta-\tau; (1-\beta)b-y, -\mu_-) h(t-\eta; -x, -\mu_-) d\eta d\tau db \\
&= 2(1-\beta) e^{2\mu_-y} \int_0^\infty \int_0^t h(\tau; (1+\beta)b, \mu_+) \int_0^{t-\tau} h(t-\tau-u; (1-\beta)b-y, -\mu_-) h(u; -x, -\mu_-) du d\tau db \\
&= 2(1-\beta) e^{2\mu_-y} \int_0^\infty \int_0^t h(\tau; (1+\beta)b, \mu_+) h(t-\tau; (1-\beta)b-y-x, -\mu_-) d\tau db.
\end{aligned}$$

We change the order of τ and η in the second equality and use changing variable $u = t - \eta$ in the third equality. So,

$$\begin{aligned}
p(x, y, t) &= \phi(y-x; \mu_-) - e^{2\mu_-y} \phi(-x-y; \mu_-) \\
&\quad + 2(1-\beta) e^{2\mu_-y} \int_0^\infty \int_0^t h(\tau; (1+\beta)b, \mu_+) h(t-\tau; (1-\beta)b-y-x, -\mu_-) d\tau db.
\end{aligned}$$

□

1.10 Remarks on the joint density in Gairat and Shcherbakov

[48]

Based on equation (2.3) in their paper

In [48] the following joint density of the refracted skew Brownian motion X_T , its last visit to

the origin τ , occupation time V , and local times $L_T^{(0)}(X)$ is given in (2.3), where T is time.

For $0 \leq v \leq t \leq T, l \geq 0$,

$$\phi_T(t, v, x, l) = \begin{cases} 2ph'(v, lp)h'(t-v, lq)h'(T-t, x)e^{-\frac{m_1^2 v}{2} - \frac{m_2^2 (T-v)}{2} - l(m_1 p - qm_2) + m_1 x}, & x > 0, \\ 2qh'(v, lp)h'(t-v, lq)h'(T-t, x)e^{-\frac{m_1^2 v}{2} - \frac{m_2^2 (T-v)}{2} - l(m_1 p - qm_2) + m_2 x}, & x < 0, \end{cases}$$

where $h'(s; y) = \frac{|y|}{\sqrt{2\pi s^3}} e^{-\frac{y^2}{2s}}$, $p = 1 - q$ and m_1 and m_2 correspond to μ_- and μ_+ respectively in our notation. We can write function h in (17) as $h(s; y, \mu) = h'(s; y)e^{-\frac{\mu^2 s}{2}} e^{-\mu y}$.

Now, let's find the marginal distribution of X_T , for $x > 0$:

$$\begin{aligned} \mathbb{P}_T(x) &= \int_0^\infty \int_0^T \int_0^t \phi_T(t, v, x, l) dv dt dl \\ &= \int_0^\infty \int_0^T \int_0^t 2ph'(v, lp)h'(t-v, lq)h'(T-t, x)e^{-\frac{m_1^2 v}{2} - \frac{m_2^2 (T-v)}{2} - l(m_1 p - qm_2) + m_1 x} dv dt dl \\ &= 2pe^{m_1 x} \int_0^\infty \int_0^T \int_0^t h'(v, lp)e^{-\frac{m_1^2 v}{2}} e^{-m_1 lp} h'(t-v, lq)h'(T-t, x)e^{-\frac{m_2^2 (T-t+v)}{2} + m_2 lq} dv dt dl \\ &= 2pe^{m_1 x} \int_0^\infty \int_0^T \int_0^t h'(v, lp)e^{-\frac{m_1^2 v}{2}} e^{-m_1 lp} h'(t-v, lq)e^{-\frac{m_2^2 (t-v)}{2}} e^{m_2 lq} h'(T-t, x)e^{-\frac{m_2^2 (T-t)}{2}} dv dt dl \\ &= 2pe^{m_1 x} \int_0^\infty \int_0^T \int_0^t h(v; lp, m_1)h(t-v; lq, m_2)e^{2m_2 lq} h(T-t; x, m_2)e^{m_2 x} dv dt dl. \end{aligned}$$

Now we change the order of v and t in the triple integration:

$$\begin{aligned} \mathbb{P}_T(x) &= 2pe^{m_1 x} \int_0^\infty \int_0^T \int_v^T h(v; lp, m_1)h(t-v; lq, m_2)e^{2m_2 lq} h(T-t; x, m_2)e^{m_2 x} dt dv dl \\ &= 2pe^{m_1 x} \int_0^\infty \int_0^T h(v; lp, m_1)e^{2m_2 lq} e^{m_2 x} \int_v^T h(t-v; lq, m_2)h(T-t; x, m_2) dt dv dl. \end{aligned}$$

Changing variable $u = T - t$ we have

$$\begin{aligned}
\mathbb{P}_T(x) &= 2pe^{m_1x} \int_0^\infty \int_0^T h(v; lp, m_1) e^{2m_2lq} e^{m_2x} \int_0^{T-v} h(T-v-u; lq, m_2) h(u; x, m_2) du dv dl \\
&= 2pe^{m_1x} \int_0^\infty \int_0^T h(v; lp, m_1) e^{2m_2lq} e^{m_2x} h(T-v; lq+x, m_2) dv dl \\
&= 2pe^{m_1x} \int_0^\infty \int_0^T h(v; lp, m_1) h(T-v; lq+x, -m_2) e^{-m_2x} dv dl \\
&= 2pe^{(m_1-m_2)x} \int_0^\infty \int_0^T h(v; lp, m_1) h(T-v; lq+x, -m_2) dv dl,
\end{aligned}$$

which is not equal to probability density in Theorem (1.6).

for $x < 0$:

$$\begin{aligned}
\mathbb{P}_T(x) &= \int_0^\infty \int_0^T \int_v^T \phi_T(t, v, x, l) dt dv dl \\
&= \int_0^\infty \int_0^T \int_v^T 2qh'(v, lp) h'(t-v, lq) h'(T-t, x) e^{-\frac{m_1^2v}{2} - \frac{m_2^2(T-v)}{2} - l(m_1p - qm_2) + m_2x} dt dv dl \\
&= 2q \int_0^\infty \int_0^T \int_v^T h'(v, lp) e^{-\frac{m_1^2v}{2}} e^{-m_1lp} h'(t-v, lq) h'(T-t, x) e^{-\frac{m_2^2(T-t+t-v)}{2} + m_2lq + m_2x} dt dv dl \\
&= 2q \int_0^\infty \int_0^T \int_v^T h'(v, lp) e^{-\frac{m_1^2v}{2}} e^{-m_1lp} h'(t-v, lq) e^{-\frac{m_2^2(t-v)}{2}} e^{m_2lq} h'(T-t, x) e^{-\frac{m_2^2(T-t)}{2}} e^{m_2x} dt dv dl \\
&= 2q \int_0^\infty \int_0^T \int_v^T h(v; lp, m_1) h(t-v; lq, m_2) e^{2m_2lq} h(T-t; x, m_2) e^{2m_2x} dt dv dl \\
&= 2q \int_0^\infty \int_0^T h(v; lp, m_1) e^{2m_2lq} \int_v^T h(t-v; lq, m_2) h(T-t; -x, m_2) dt dv dl.
\end{aligned}$$

By Changing of variable $u = T - t$:

$$\begin{aligned}
\mathbb{P}_T(x) &= 2q \int_0^\infty \int_0^T h(v; lp, m_1) e^{2m_2lq} \int_0^{T-v} h(T-v-u; lq, m_2) h(u; -x, m_2) du dv dl \\
&= 2q \int_0^\infty \int_0^T h(v; lp, m_1) e^{2m_2lq} h(T-v; lq-x, m_2) dv dl \\
&= 2q \int_0^\infty \int_0^T h(v; lp, m_1) e^{2m_2(lq-x)} e^{2m_2x} h(T-v; lq-x, m_2) dv dl \\
&= 2qe^{2m_2x} \int_0^\infty \int_0^T h(v; lp, m_1) h(T-v; lq-x, -m_2) dv dl,
\end{aligned}$$

which is equal to probability density in Theorem (1.6).

Based on equation (2.4) in their paper

Based on the same paper, the joint density of the refracted skew Brownian motion (X_T) , its last visit to the origin (τ) , total occupation time (U) , and local times $(L_T^{(0)}(X))$ is given by equation (2.4):

$$\varphi_T(t, v, x, l) = \begin{cases} 2ph'(u+t-T, lp)h'(T-u, lq)h'(T-t, x)e^{-\frac{m_1^2 u}{2}-\frac{m_2^2(T-u)}{2}-l(m_1 p-qm_2)+m_1 x}, & x > 0, \\ l > 0, t \leq T, T-t \leq u \leq T, \\ 2qh'(u, lp)h'(t-u, lq)h'(T-t, x)e^{-\frac{m_1^2 u}{2}-\frac{m_2^2(T-u)}{2}-l(m_1 p-qm_2)+m_2 x}, & x < 0, \\ l > 0, 0 \leq u \leq t \leq T. \end{cases}$$

Now, let's find the marginal distribution of X_T , for $x > 0$:

$$\begin{aligned} \mathbb{P}_T(x) &= \int_0^\infty \int_0^T \int_{T-t}^T \varphi_T(t, v, x, l) du dt dl \\ &= \int_0^\infty \int_0^T \int_{T-t}^T 2ph'(u+t-T, lp)h'(T-u, lq)h'(T-t, x)e^{-\frac{m_1^2 u}{2}-\frac{m_2^2(T-u)}{2}-l(m_1 p-qm_2)+m_1 x} du dt dl \\ &= 2pe^{m_1 x} \int_0^\infty \int_0^T \int_{T-t}^T h'(u+t-T, lp)h'(T-u, lq)h'(T-t, x)e^{-\frac{m_1^2 u}{2}-\frac{m_2^2(T-u)}{2}-l(m_1 p-qm_2)} du dt dl \\ &= 2pe^{m_1 x} \int_0^\infty \int_0^T \int_{T-t}^T h'(u+t-T, lp)e^{-\frac{m_1^2}{2}(u+t-T)}e^{-m_1 lp}h'(T-u, lq)e^{-\frac{m_2^2(T-u)}{2}+lqm_2}h'(T-t, x) \\ &\quad \times e^{-\frac{m_1^2(T-t)}{2}} du dt dl \\ &= 2pe^{m_1 x} \int_0^\infty \int_0^T \int_{T-t}^T h(u+t-T; lp, m_1)h(T-u; lq, m_2)e^{2lqm_2}h'(T-t; x, m_1)e^{m_1 x} du dt dl. \end{aligned}$$

Now we change the order of u and t in the triple integration:

$$\mathbb{P}_T(x) = 2pe^{2m_1 x} \int_0^\infty \int_0^T \int_{T-u}^T h(u+t-T; lp, m_1)h(T-u; lq, m_2)e^{2lqm_2}h'(T-t; x, m_1) dt du dl.$$

Using the change of variable $v = T - t$,

$$\begin{aligned}
\mathbb{P}_T(x) &= 2pe^{2m_1x} \int_0^\infty \int_0^T \int_0^u h(u-v; lp, m_1) h(T-u; lq, m_2) e^{2lqm_2} h'(v; x, m_1) dv du dl \\
&= 2pe^{2m_1x} \int_0^\infty \int_0^T h(T-u; lq, -m_2) \int_0^u h(u-v; lp, m_1) h'(v; x, m_1) dv du dl \\
&= 2pe^{2m_1x} \int_0^\infty \int_0^T h(T-u; lq, -m_2) h(u; lp+x, m_1) du dl,
\end{aligned}$$

which is equal to probability density in Theorem (1.6).

Now, let's find the marginal distribution of X_T , for $x < 0$:

$$\begin{aligned}
\mathbb{P}_T(x) &= \int_0^\infty \int_0^T \int_u^T \varphi_T(t, v, x, l) du dt dl \\
&= \int_0^\infty \int_0^T \int_u^T 2qh'(u, lp) h'(t-u, lq) h'(T-t, x) e^{-\frac{m_1^2 u}{2} - \frac{m_2^2 (T-u)}{2} - l(m_1 p - qm_2) + m_2 x} dt du dl \\
&= 2q \int_0^\infty \int_0^T \int_u^T h'(u, lp) e^{-\frac{m_1^2 u}{2} - m_1 lp} h'(t-u, lq) e^{-\frac{m_2^2 (t-u)}{2} + m_2 lq} h'(T-t, x) e^{-\frac{m_2^2 (T-t)}{2} + m_2 x} dt du dl \\
&= 2q \int_0^\infty \int_0^T \int_u^T h(u; lp, m_1) h(t-u; lq, m_2) e^{2m_2 lq} h(T-t; x, m_2) e^{2m_2 x} dt du dl.
\end{aligned}$$

Using the change of variable $v = T - t$,

$$\begin{aligned}
\mathbb{P}_T(x) &= 2q \int_0^\infty \int_0^T \int_0^{T-u} h(u; lp, m_1) h(T-u-v; lq, m_2) e^{2m_2 lq} h(v; x, m_2) e^{2m_2 x} dv du dl \\
&= 2q \int_0^\infty \int_0^T h(u; lp, m_1) e^{2m_2 lq} \int_0^{T-u} h(T-u-v; lq, m_2) h(v; -x, m_2) dv du dl \\
&= 2q \int_0^\infty \int_0^T h(u; lp, m_1) e^{2m_2 lq} h(T-u; lq-x, m_2) du dl \\
&= 2qe^{2m_2 x} \int_0^\infty \int_0^T h(u; lp, m_1) h(T-u; lq-x, m_2) e^{2m_2 lq} e^{-2m_2 x} du dl \\
&= 2qe^{2m_2 x} \int_0^\infty \int_0^T h(u; lp, m_1) h(T-u; lq-x, -m_2) du dl,
\end{aligned}$$

which is equal to probability density in Theorem (1.6).

Chapter 2

Greenhouse-Gas-Emission-Aware Portfolio Optimization with Deep Reinforcement Learning

2.1 Introduction

According to the Sixth Assessment Report by the Intergovernmental Panel on Climate Change [65], the average annual global greenhouse gas (GHG) emissions over the last decade (2010–2019) were higher than in any previous decade across all categories of GHGs. This level of GHG emissions, coupled with concerns about future environmental impacts, has prompted governments and international organizations to implement policies aimed at reducing such emissions. Key examples include the Paris Agreement [104] among UN members and the U.S. long-term climate strategy [105]. Additionally, the growing awareness of environmental challenges has accelerated the integration of Environmental, Social, and Governance (ESG) criteria into investment and corporate decision-making processes. This shift underscores the importance of sustainable and responsible practices in mitigating the impacts of climate change. Green portfolio optimization has emerged as a critical research area, driven by an increasing emphasis on sustainable investing. As of early 2022,

the *Global Sustainable Investment Review* reported over 30 trillion USD invested with specific ESG objectives [4]. Additionally, the 2017-2018 annual report from the Principles for Responsible Investment (PRI), an initiative endorsed by the United Nations, indicated that its signatories oversee nearly \$90 trillion in assets [99]. There is a serious determination to reduce GHG emissions globally and within the United States. Laws, regulations, and public concerns are pushing institutional investors towards socially responsible investing.

In recent years, investor preferences have also shifted towards assets that are viewed as more environmentally sustainable and potentially less vulnerable to the negative impacts of climate change. By example, Pástor et al. [95] show that green stocks are increasingly attracting investors, often at the expense of stocks from carbon-intensive companies. Socially Responsible Investing (SRI) is an investment strategy that incorporates ESG considerations alongside financial returns. Rather than focusing solely on profits, SRI investors aim to align their investments with their personal values and contribute to a more sustainable world. This alignment is achieved by integrating ESG factors into various stages of the investment process, including research, analysis, and security selection [17, 36, 95, 9]. As noted by Widyawati [113], tools and empirical knowledge about how to effectively implement SRI quantitative financial models are limited, which hinders their adoption in practice.

This paper contributes to the development of quantitative portfolio management methods compatible with SRI considerations. Our main objective is to develop a novel portfolio optimization scheme which provides a return-risk-GHG emissions tradeoff compatible with the investor's preferences. More precisely, the approach followed is referred to as mean-var-E,¹ where we maximize the portfolio's expected wealth while simultaneously minimizing its risk and GHG emissions. Our approach extends several works from the SRI stream of literature who consider static optimization [96, 7, 79], as we consider multi-period optimization.² In such framework, the allocation procedure (i) suitably identifies assets with most favorable financial-environmental performance tradeoff, (ii) reacts to time-varying conditions of the market and exhibits improved adaptability, and (iii) remains

¹The letter "E" represents the environmental component of ESG.

²We extend the work of van Staden et al. [107] by incorporating GHG emissions into the objective function.

optimal on a long-run perspective by considering statistical interactions between returns of the various periods. Our framework makes use of well-established econometric models to perform realistic projections of returns and GHG emissions for the various asset classes. In particular, reliance on stochastic modeling acknowledges the inherent uncertainty and variability in GHG emissions, leading to more realistic portfolio strategies that can better accommodate real-world environmental impact fluctuations. The mean-var-E portfolio optimization in such setting is a very challenging task due to its complex high-dimensional nature; no closed-form solution can be obtained. Recent developments in reinforcement learning, a branch of machine learning incorporating methods to solve sequential decision making problems, are leveraged in this study to tackle the multi-period mean-var-E problem.

Our approach directly considers GHG emissions and contrasts several other works which instead use ESG scores [60, 51, 27]. Whereas emissions are an objective quantity that can be measured, ESG scores are more subjective in nature and exhibit several issues related to the lack of standardized approach for their construction. This can result in inconsistent ESG portfolio construction [113, 16]. ESG scores are as such also more susceptible to "greenwashing" [31]. By incorporating emission intensities rather than relying on ESG scores, our approach mitigates the risks of inconsistency and greenwashing, improving the alignment of the portfolio allocations towards environmental sustainability objectives. The scarcity of emissions data makes their projection in the future challenging. We acknowledge that this entails non-negligible model risk. Nevertheless, we settle for a simple a reasonable model; tackling the climate crisis is urgent, and it is unreasonable to wait to get sufficient data and obtain a precise stochastic model for emissions before starting de-carbonating investment portfolios. Extensions of our model integrating either robustification of decisions [67] or model ambiguity [88] could be contemplated as interesting future work.

We test our approach within a sectorial allocation framework encompassing the various sectors of the S&P 500, focusing explicitly on scope 1 and scope 2 GHG emissions intensities. These scopes capture direct emissions from a company's operations and indirect emissions from the generation of purchased energy, respectively, providing a measure of a company's environmental impact.

Scope 3 emissions were not considered since data for such scope is often deemed less reliable. This sector-specific analysis allows us to assess the real-world applicability of our optimization model and its potential to drive meaningful reductions in GHG emissions within portfolios of environmentally-aware investors.

Numerical experiments reveal that our approach outperform considered benchmarks, which are the equally-weighted, the optimal static and the myopic dynamic approaches. In some cases, the RL approach dominates the benchmarks by providing simultaneously higher expected return, lower risk and lower GHG emissions over the full investment horizon. The good performance of our approach can be explained by (i) the ability of our method to identify assets with favorable tradeoffs between financial and environmental performance that are aligned with the investor’s preferences, (ii) the ability to react to fluctuations in market conditions, and (iii) multi-stage optimality which integrates interactions between all portfolio rebalancing decisions applied.

The rest of the paper is organized as follows: Section 2.2 provides a more extensive review of the SRI literature, highlighting key studies related to our work. Section 2.3 presents the mean-var-E portfolio optimization problem along with the proposed solution approach based on reinforcement learning. Section 2.4 describes econometric models considered to represent the dynamics of asset returns and GHG emissions, along with their calibration to the various sectors of the S&P 500 index which serve as assets in the model. Numerical experiments showcasing the application of our model are provided in Section 2.5. Section ?? concludes. The code used to produce numerical results of this paper is available on Github.³

2.2 Literature review

Research on SRI portfolio allocation has been ongoing for years. Kempf and Osthoff [73] study the influence of socially responsible criteria on stock portfolio performance using three different screens: negative, positive, and best-in-class. Ballesterio et al. [13] address a large-scale

³See <https://github.com/zaniara3/DCC-EGARCH-Simulation> and <https://github.com/zaniara3/green-portfolio-optimization>.

portfolio selection problem using classical utility theory under uncertainty with an SRI perspective, dividing the asset universe into two subsets: ethical assets and non-ethical assets. Dorfleitner et al. [38] extend the classic mean-variance (Markowitz) portfolio theory by including a social return component. This notion can be implemented using environmental, social, and governance rating scores (see Dorfleitner and Utz [37] for more details). Hirschberger et al. [60] and Utz et al. [106] extend the Markowitz portfolio selection to a tri-criterion approach, using the ESG score of firms as one of the criteria.

Building on this perspective, Dorfleitner and Nguyen [36] present two approaches to model the dependency of optimal socially responsible portfolios on available investment volume in an expected utility context where mean-variance optimization with an additional social dimension is applied.⁴ Geczy et al. [52] construct optimal portfolios using a multifactor expected return framework based on the arbitrage pricing theory. They optimize returns while constraining tracking error. They demonstrate that including ESG criteria can potentially enhance stockholder returns while keeping risk constant under certain assumptions. In another study, Alessandrini and Jondeau [2] propose optimal strategies for building ESG portfolios, incorporating benchmark weights, industry classifications, and factor exposures. They argue that investing based on screening would generate undesirable regional, sectorial, and risk factor exposures. A year later, this was confirmed when Amon et al. [5] analyzed several passive asset allocation strategies based on ESG-weighting. Chen et al. [27] suggest a three-stage approach for creating SRI portfolios using a data envelopment analysis (DEA) model. They demonstrate that the produced SRI portfolio may outperform common investment methods in various ways while maintaining investment and social value consistency. Wu et al. [114] develop a hybrid ESG portfolio selection model that combines multi-criteria decision-making and multi-objective optimization.

In addition to these optimization models, Lauria et al. [76] propose a unified framework for integrating ESG ratings into dynamic pricing theory and determining optimal reward-risk-ESG portfolio allocation. Similarly, Liagkouras et al. [83] propose an algorithm that first executes

⁴They use social return in the model, see [37] for more details.

a screening to remove stocks from the investment universe that do not follow the given ESG restrictions, and then runs the portfolio optimization on the ESG-compliant universe. Abate et al. [1] integrate ESG scores with both mean-variance and mean-downside risk frameworks, compare their performances, and address whether a sustainable investor will experience a reduction or increase in portfolio risk-adjusted returns compared to a traditional investor. Liu et al. [84] introduce a novel portfolio optimization approach for the energy sector, incorporating behavioral aspects like regret theory and using cross-efficiency analysis from DEA. They investigate the ESG and environmental performance of the proposed portfolios. Xidonas and Essner [115] consider a multi-objective minimax-based portfolio optimization involving the maximization of ESG risk performance.

Furthermore, Anquetin et al. [7] incorporate carbon emissions into a static mean-variance portfolio optimization as a third dimension, demonstrating that with adequate carbon constraints, reducing emission intensities by at least half is feasible while incurring little to no reduction in the Sharpe ratio.

Additional works focus on dynamic approaches to socially responsible investing. Esparcia et al. [44] present a dynamic portfolio optimization model that minimizes risk exposure and evaluates the performance of green and brown portfolios based on the Sharpe ratio, as well as downside and upside risks. The results demonstrate that green portfolios consistently outperform brown portfolios and passive strategies. Additionally, several studies have utilized reinforcement learning techniques to dynamically determine optimal strategies for SRI portfolios. For instance, in Vo et al. [109], reinforcement learning is employed to develop a deep responsible investment portfolio model, which incorporates ESG scores into the mean-variance model. Furthermore, Maree and Omlin [87] compare a multi-agent deep deterministic policy gradient system with a system that uses a genetic algorithm to optimize the parameters of the deep neural networks of three agents respectively concerned with profit, risk and ESG criteria.

Coqueret [31] identifies growing trends indicating that investors are increasingly allocating resources to green enterprises. This shift suggests that sustainable enterprises are perceived as safer investments. This changing investment behavior reflects an increased understanding of sustainability

as a crucial component of financial performance. The Coqueret [31] study emphasizes the difficulty of evaluating whether SRI improves or diminishes portfolio performance. This difficulty depends on the ESG components that play distinct roles, the regional context, specific market conditions and the time-varying nature of the market.

2.3 General Framework

This section introduces a general framework for GHG-emission-averse portfolio optimization. We consider a discrete-time economy defined on the time set $\{0, 1, \dots, T\}$ and consisting of N risky assets. The uncertainty is modeled with the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ represents the physical measure and operators $\mathbb{E}$ and $\mathbb{V}ar$ denote the expectation and the variance, respectively. The return from time $t - 1$ to t of the j -th risky asset, $j = 1, \dots, N$, is denoted by $r_{t,j} = \frac{p_{t,j} - p_{t-1,j}}{p_{t-1,j}}$, where $p_{t,j}$ is the asset price at time t . The risky asset returns vector is $\mathbf{r}_t = (r_{t,1}, \dots, r_{t,N})^\top$. Denote also by $\mathbf{g}_t = (g_{t,1}, \dots, g_{t,N})^\top$ the periodic greenhouse gas intensity of risky assets at time t . Each $g_{t,j}$ is a ratio of the greenhouse gas emissions to the market capitalization:

$$g_{t,j} = \frac{GHG_{t,j}}{MKT_{t-1,j}}, \quad j = 1, 2, \dots, N, \quad (27)$$

where $GHG_{t,j}$ is the greenhouse gas emission (in kilo tons CO₂eq, the “eq” standing for equivalent) of the risky asset j for period $(t - 1, t]$ and $MKT_{t-1,j}$ is the market capitalization (in million of \$) of the asset j at time $t - 1$. We refer to the set $\mathbf{Y} = \{\mathbf{Y}_t = (\mathbf{r}_t^\top, \mathbf{g}_t^\top), t = 1, \dots, T\}$ as the path of (joint) asset returns and greenhouse gas intensities over the investment time horizon T .

At each time t , the investor observes the vector $\mathbf{x}_t$, which represents the information taken into account to determine their asset allocation decision. The composition of $\mathbf{x}_t$ in our setting is detailed subsequently.

Given $\mathbf{x}_t$, the weight $\omega_{t,j}(\mathbf{x}_t)$ is the fraction of wealth available at time t invested in the asset j during $(t, t + 1]$, $j = 1, \dots, N$. The vector containing all such weights is $\boldsymbol{\omega}_t(\mathbf{x}_t) = (\omega_{t,1}(\mathbf{x}_t), \dots, \omega_{t,N}(\mathbf{x}_t))^\top$. Weights must always sum to one, i.e. $\sum_{j=1}^N \omega_{t,j} = 1$. Moreover, no

short selling is considered: $w_{t,j} \geq 0$. For notation ease, we omit the parentheses in $\omega_t(\mathbf{x}_t)$ and $\omega_{t,j}(\mathbf{x}_t)$ and we simply write ω_t and $\omega_{t,j}$. We use $\Omega = \{\omega_t, t = 1, \dots, T\}$ to designate the allocation sequence.

The portfolio cumulative wealth and cumulative greenhouse gas emission at time $t + 1$ are given by

$$W_{t+1}(\Omega, \mathbf{Y}) = W_0(\Omega, \mathbf{Y}) \prod_{s=0}^t \omega_s^\top (\mathbf{r}_{s+1} + 1) = W_t(\Omega, \mathbf{Y}) \omega_t^\top (\mathbf{r}_{t+1} + 1), \quad (28)$$

$$G_{t+1}(\Omega, \mathbf{Y}) = G_t(\Omega, \mathbf{Y}) + \omega_t^\top \mathbf{g}_{t+1} W_t(\Omega, \mathbf{Y}), \quad (29)$$

where the initial wealth $W_0(\Omega, \mathbf{Y})$ and the initial greenhouse gas emission $G_0(\Omega, \mathbf{Y})$ are supposed, without loss of generality, to be respectively equal to 100 and 0.

The objective of the portfolio manager is to find the optimal allocation sequence Ω^* which offers an optimal tradeoff between portfolio expected return, financial risk and carbon emissions. As such we propose to use the following objection function:

$$\mathcal{O}(W_T(\Omega, \mathbf{Y}), G_T(\Omega, \mathbf{Y})) = \mathbb{E}[W_T(\Omega, \mathbf{Y})] - \rho \mathbb{V}ar(W_T(\Omega, \mathbf{Y})) - \gamma \mathbb{E}[G_T(\Omega, \mathbf{Y})], \quad (30)$$

where ρ is the financial risk aversion coefficient and γ is the coefficient reflecting aversion to greenhouse gas emissions. The multi-stage portfolio optimization problem is

$$\arg \max_{\Omega} \mathcal{O}(W_T(\Omega, \mathbf{Y})) \text{ under constraints } \begin{cases} \mathbf{1}^\top \omega_t = 1, \forall t = 1, \dots, T, \\ \omega_{t,j} \geq 0, \forall t = 1, \dots, T, \forall j = 1, \dots, N. \end{cases} \quad (31)$$

Note that the constraints of (i) no short-selling and (ii) no leverage could be adjusted without difficulty to treat different constraint formulations.

We choose a neural network to represent the function mapping inputs $\mathbf{x}_t$ into allocation actions

$\omega_t(\mathbf{x}_t)$. We use a fully-connected, feed-forward neural network f with parameter vector $\boldsymbol{\theta}$:

$$f(\mathbf{x}_t, \boldsymbol{\theta}) = \omega_t. \quad (32)$$

To estimate the objective function (30), Monte Carlo simulation is applied. We first simulate n paths $\mathbf{Y}^{(i)}$, $i = 1, \dots, n$, each of which consists in a realization of joint asset returns $(\mathbf{r}_t)_{t=1, \dots, T}$ and periodic greenhouse gas intensity $(\mathbf{g}_t)_{t=1, \dots, T}$. Then, for any set of policy parameters $\boldsymbol{\theta}$, the associated portfolio wealth and cumulative greenhouse gas emission along path $\mathbf{Y}^{(i)}$ are obtained using equations (28) and (29), where weights for all time steps are calculated with (32). The estimate of the objective function (30) is given by

$$\mathcal{J}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \left[W_T(\boldsymbol{\theta}, \mathbf{Y}^{(i)}) - \rho \left(W_T(\boldsymbol{\theta}, \mathbf{Y}^{(i)}) - \frac{1}{n} \sum_{j=1}^n W_T(\boldsymbol{\theta}, \mathbf{Y}^{(j)}) \right)^2 - \gamma G_T(\boldsymbol{\theta}, \mathbf{Y}^{(i)}) \right]. \quad (33)$$

The optimization problem (31) is therefore approximated by

$$\arg \max_{\boldsymbol{\theta}} \mathcal{J}(\boldsymbol{\theta}), \quad (34)$$

which is itself solved using stochastic gradient ascent (SGA) techniques. The parameter set is sequentially refined to produce a sequence of estimates $\{\boldsymbol{\theta}^{(j)}\}_{j \geq 1}$ through

$$\boldsymbol{\theta}^{(j+1)} = \boldsymbol{\theta}^{(j)} + \eta_j \nabla_{\boldsymbol{\theta}} \hat{\mathcal{J}}(\boldsymbol{\theta}^{(j)}), \quad (35)$$

where learning step sizes $\{\eta_j\}$ are obtained with the Adam algorithm, and the estimate of the objective function gradient $\nabla_{\boldsymbol{\theta}} \hat{\mathcal{J}}(\boldsymbol{\theta}^{(j)})$ is obtained through automatic differentiation tools provided by `Pytorch` using a batch of paths of size n_{batch} . Parameters of the neural network are initialized with the Kaiming uniform initialization [59], which gives the initial value of the sequence $\boldsymbol{\theta}^{(0)}$. We also apply batch normalization on the input layer to accelerate neural network training.

2.4 Econometric model for asset returns and emissions

This section describes the considered econometric model that acts as the environment for the reinforcement learning algorithm and allows simulating paths of state variables. We consider an investor interested in sectorial allocation within the universe of the S&P 500, with sectors categorized according to the Global Industry Classification Standard (GICS). Dynamics of the sectorial asset returns and greenhouse gas emission intensities in the model are hereby described.

2.4.1 DCC-EGARCH model for asset returns

Model specification

Log-returns of the various sectors are assumed to follow a Dynamic Conditional Correlation Exponential Generalized Autoregressive Conditional Heteroskedasticity (DCC-EGARCH) model. Such model allows for time-varying volatilities and correlations which are well-recognized stylized facts of financial returns. Following [93] and [43], given the information set available at time $t - 1$, the conditional distribution of the time- t log-returns is multivariate normal distribution with mean $\boldsymbol{\mu} + \mathbf{D}_t^2 \boldsymbol{\kappa}$ and variance-covariance matrix $\mathbf{H}_t$, whose dynamics is given by

$$\log(1 + \mathbf{r}_t) = \boldsymbol{\mu} + \mathbf{D}_t^2 \boldsymbol{\kappa} + \mathbf{a}_t \quad (36)$$

$$\mathbf{a}_t = \mathbf{H}_t^{1/2} \mathbf{z}_t, \quad (37)$$

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t, \quad (38)$$

$$\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N), \quad (39)$$

where $\mathbf{D}_t$ is an $N \times N$ diagonal matrix of time-varying volatilities from the univariate EGARCH model, $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{H}_t^{1/2}$ is the lower triangular matrix of the Cholesky decomposition of $\mathbf{H}_t$. The j^{th} element on the diagonal of both $\mathbf{H}_t$ and $\mathbf{D}_t^2$ is the time-varying

conditional variance of asset j denoted by $h_{t,j}$. Its EGARCH dynamics is

$$\log(h_{t,j}) = \omega_j + \alpha_j(|\epsilon_{t-1,j}| - \sqrt{2/\pi}) + \gamma_j \epsilon_{t-1,j} + \beta_j \log(h_{t-1,j}).$$

Also, $\mathbf{R}_t$ is the conditional correlation matrix of the standardized disturbances $\boldsymbol{\epsilon}_t$ defined as

$$\boldsymbol{\epsilon}_t = \mathbf{D}_t^{-1} \mathbf{a}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t).$$

In the DCC model, the time-varying correlation matrix $\mathbf{R}_t$ evolves according to

$$\mathbf{R}_t = \mathbf{Q}_t^{\star-1} \mathbf{Q}_t \mathbf{Q}_t^{\star-1},$$

where $\mathbf{Q}_t^{\star} = \text{diag}(\sqrt{q_{t,11}}, \dots, \sqrt{q_{t,NN}})$ and $\mathbf{Q}_t = [q_{t,ij}]_{i,j=1}^N$ is a symmetric positive-definite matrix following the dynamics

$$\mathbf{Q}_t = (1 - \alpha^{DCC} - \beta^{DCC}) \overline{\mathbf{Q}} + \alpha^{DCC} \boldsymbol{\epsilon}_{t-1} \boldsymbol{\epsilon}_{t-1}^{\top} + \beta^{DCC} \mathbf{Q}_{t-1}, \quad (40)$$

with $\overline{\mathbf{Q}} = \text{Cov}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{\top}] = \mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^{\top}]$ being the unconditional covariance of the standardized residuals.

Parameters α^{DCC} and β^{DCC} are non-negative with $\alpha^{DCC} + \beta^{DCC} < 1$.

Data and estimation

The assets considered are these of the eleven sectors of the S&P 500 index, namely Utilities (UTL), Energy (ENG), Industrials (IND), Materials (MAT), Consumer Discretionary (CDI), Consumer Staples (CST), Communication Services (CSV), Information Technology (ITC), Health Care (HEC), Real Estate (RES), and Financials (FIN). Our data consists of monthly returns over the period November 2001 to December 2022 (254 months). The data is downloaded from CRSP. Returns are computed using the closing price on the last day of each month. Univariate descriptive statistics of the log-returns $\log(1 + r_{t,j})$ are shown in Table 2.1, whereas historical correlations are

shown in Table 2.2.

Table 2.1: Summary statistics for monthly returns of the S&P 500 sectors.

Sectors	Mean	StD	Skewness	Kurtosis	Min	Q1 25%	Q3 75%	Max
ENG	0.0069	0.0733	-0.678	5.481	-0.4276	-0.0277	0.0468	0.2606
HEC	0.0071	0.0394	-0.435	0.613	-0.1335	-0.0175	0.0309	0.1191
IND	0.0067	0.0543	-0.778	2.380	-0.2130	-0.0198	0.0379	0.1643
MAT	0.0073	0.0592	-0.573	1.713	-0.2494	-0.0240	0.0435	0.1632
FIN	0.0040	0.0639	-1.123	4.226	-0.3053	-0.0223	0.0397	0.2018
CST	0.0073	0.0335	-0.580	0.914	-0.1156	-0.0116	0.0294	0.0979
UTL	0.0065	0.0442	-0.864	1.266	-0.1506	-0.0167	0.0353	0.1149
ITC	0.0086	0.0604	-0.488	0.963	-0.1959	-0.0232	0.0475	0.2013
RES	0.0044	0.0648	-1.424	8.826	-0.3869	-0.0226	0.0411	0.2984
CSV	0.0031	0.0555	-0.103	2.645	-0.1698	-0.0253	0.0382	0.2848
CDI	0.0073	0.0546	-0.345	1.664	-0.2131	-0.0157	0.0413	0.1869

Table 2.2: Historical correlations between pairs of monthly returns from sectors of the S&P 500.

Sectors	ENG	HEC	IND	MAT	FIN	CST	UTL	ITC	RES	CSV	CDI
ENG	1										
HEC	0.420	1									
IND	0.630	0.696	1								
MAT	0.651	0.652	0.864	1							
FIN	0.563	0.646	0.855	0.758	1						
CST	0.414	0.681	0.689	0.608	0.618	1					
UTL	0.355	0.461	0.437	0.410	0.360	0.583	1				
ITC	0.461	0.637	0.759	0.743	0.675	0.557	0.397	1			
RES	0.383	0.561	0.689	0.649	0.716	0.594	0.508	0.566	1		
CSV	0.439	0.533	0.617	0.558	0.523	0.559	0.496	0.648	0.440	1	
CDI	0.497	0.671	0.834	0.790	0.776	0.639	0.395	0.821	0.694	0.623	1

Returns from the eleven sectors are highly correlated, with the smallest correlation being 0.355 (Energy vs Utilities) and the highest being 0.864 (Industrial vs Materials).

The loglikelihood function for the DCC-EGARCH model applied to the entire set of returns is expressed as

$$\mathcal{L}(\boldsymbol{\vartheta}) = -\frac{1}{2} \sum_{t=1}^T \left(N \log(2\pi) + 2 \log(|\mathbf{D}_t|) + \log(|\mathbf{R}_t|) + \boldsymbol{\epsilon}_t' \mathbf{R}_t^{-1} \boldsymbol{\epsilon}_t \right), \quad (41)$$

where $\boldsymbol{\vartheta}$ denotes the model parameters. These parameters are partitioned into two distinct groups: $(\boldsymbol{\phi}, \boldsymbol{\varphi}) = (\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_N, \boldsymbol{\varphi})$. Herein, $\boldsymbol{\phi}_j = (\mu_j, \kappa_j, \omega_j, \alpha_j, \gamma_j, \beta_j)$ contains the parameters of the univariate EGARCH model for the j -th sector, while $\boldsymbol{\varphi} = (\alpha^{DCC}, \beta^{DCC})$ represents the

parameters of the correlation structure in (40). Matrix $\overline{Q}$ is set to the sample covariance of standardized errors $\hat{\epsilon}_t$.

Due to the complexity of jointly optimization all parameters simultaneously, a two-stage estimation strategy is used. In the initial stage, parameters ϕ_i of the univariate EGARCH models is estimated for each sector i . In the second stage, φ is obtained by optimizing the log-likelihood (41), with parameters $\phi_i, i = 1, \dots, N$ being fixed to their estimate obtained in the first stage. In other words,⁵ first, the volatility ($\mathcal{L}_v$) is maximized:

$$\mathcal{L}_v(\phi) = -\frac{1}{2} \sum_{t=1}^T (N \log(2\pi) + 2 \log(|\mathbf{D}_t|) + \mathbf{a}_t' \mathbf{D}_t^{-1} \mathbf{D}_t^{-1} \mathbf{a}_t) ,$$

which entails estimating all univariate EGARCH processes independently. Then the DCC model likelihood ($\mathcal{L}_c$) is maximized:

$$\mathcal{L}_c(\varphi) = -\frac{1}{2} \sum_{t=1}^T (N \log(2\pi) + 2 \log(|\mathbf{D}_t|) + \log(|\mathbf{R}_t|) + \epsilon_t' \mathbf{R}_t^{-1} \epsilon_t) .$$

Since $\mathbf{D}_t$ is constant when conditioning on the parameters from step one, we can exclude the constant terms and maximize:

$$\tilde{\mathcal{L}}_c(\varphi) = -\frac{1}{2} \sum_{t=1}^T (\log(|\mathbf{R}_t|) + \epsilon_t' \mathbf{R}_t^{-1} \epsilon_t) .$$

The two-step maximum likelihood estimation of the DCC model is rigorously examined and proven, under certain conditions, to yield consistent and asymptotically normal estimators [43]. Furthermore, a comprehensive exploration conducted by Jondeau and Rockinger [70] delves into both full maximum likelihood estimation and the two-step procedure, concluding that both yield very similar results. For our empirical analysis, we employ the esteemed Broyden, Fletcher, Goldfarb, and Shanno's (BFGS) algorithm to derive the two-step maximum likelihood estimates of the multivariate DCC-EGARCH model. The estimated parameters of the DCC-EGARCH model,

⁵See [43] for details about the estimation.

along with their standard errors, are provided in Table 2.3.

Table 2.3: DCC–EGARCH parameter estimates on monthly returns of sectors of the S&P 500.

Sectors	μ_j	κ_j	ω_j	α_j	γ_j	β_j
ENG	0.002 (0.008)	0.927 (1.715)	-1.010 (0.857)	0.313 (0.109)***	-0.160 (0.123)	0.810 (0.159)***
HEC	0.004 (0.007)	1.998 (4.567)	-2.782 (1.656)*	0.103 (0.126)	-0.311 (0.110)***	0.576 (0.251)**
IND	0.005 (0.005)	0.799 (1.956)	-0.595 (0.321)*	0.175 (0.136)	-0.257 (0.066)***	0.902 (0.052)***
MAT	0.004 (0.001)***	0.979 (0.098)***	-1.252 (0.076)***	0.272 (0.075)***	-0.325 (0.064)***	0.788 (0.014)***
FIN	0.005 (0.004)	0.011 (1.307)	-0.429 (0.146)***	0.133 (0.048)**	-0.258 (0.042)***	0.927 (0.024)***
CST	-0.004 (0.009)	10.047 (8.223)	-2.489 (1.332)*	0.033 (0.135)	-0.277 (0.123)**	0.638 (0.193)***
UTL	0.016 (0.001)***	-3.183 (0.892)***	-0.460 (0.063)***	0.332 (0.092)***	0.067 (0.075)	0.913 (0.009)***
ITC	0.009 (0.014)	-0.289 (7.620)	-0.684 (0.295)**	0.306 (0.087)***	-0.158 (0.062)**	0.884 (0.048)***
RES	-0.001 (0.007)	1.271 (2.316)	-0.618 (0.367)*	0.096 (0.094)	-0.210 (0.056)***	0.891 (0.061)***
CSV	0.008 (0.009)	-1.887 (3.617)	-0.580 (0.235)**	0.151 (0.064)**	0.903 (0.051)**	0.912 (0.040)***
CDI	0.001 (0.002)	0.195 (1.313)	-1.125 (0.147)***	0.258 (0.101)**	-0.240 (0.052)***	0.811 (0.025)***

Parameter estimates, and asymptotic standard errors in parentheses. The estimates of α^{DCC} and β^{DCC} are respectively 0.035 (0.011) and 0.785 (0.155). The symbols ***(**/*) denote rejection of the null hypothesis that the parameter is equal to zero at 1% (5%/10%) significance level.

As shown in Table 2.3, all estimated EGARCH model parameters β_j are statistically positive, indicating that there is a significant GARCH effect in all sectors. Except for the Energy (ENG) and Utilities (UTL) sectors, all parameters γ_j are strongly significant and negative. This indicates that returns of the various S&P 500 sectors exhibit a strong leverage effect.

The DCC estimates of α^{DCC} and β^{DCC} in (40) are respectively 0.035 (0.011) and 0.785 (0.155). The sum of DCC parameters ($\alpha^{DCC} + \beta^{DCC}$) is much larger than zero, indicating that conditional correlations between the sectors of S&P 500 are not constant.

2.4.2 Greenhouse gas emissions time series model

Model specification

Due to the limited amount of data available (9 points per time series), we model the dynamics of the logarithm of greenhouse gas intensity of each sector with a simple random walk model with drift. The year- y log-emissions of sector j is given by

$$c_{y,j} = \bar{c}_j + c_{y-1,j} + \epsilon_{y,j}, \quad (42)$$

where $\bar{c}_j$ is a constant and $\{\epsilon_{y,j}\}$ is a white noise (also uncorrelated across sectors) with a normal distribution of mean 0 and variance $\sigma_{\epsilon,j}^2$. Once parameters for the yearly greenhouse gas log-intensities are estimated, they are converted into monthly annualized log-intensities by dividing the mean and variance of differences by 12, yielding a logarithmic intensity variation for each month t :

$$\Delta d_{t,j} = d_{t,j} - d_{t-1,j} = \bar{d}_j + \epsilon'_{t,j}, \quad t = 1, 2, \dots, T,$$

where $\bar{d}_j = \frac{\bar{c}_j}{12}$ and $\{\epsilon'_{t,j}\}$ is a white noise with zero mean and variance $\frac{\sigma_{\epsilon,j}^2}{12}$. Finally, by setting $d_{0,j} = c_{0,j}$, for all j , monthly greenhouse gas intensities are given by:

$$g_{t,j} = \frac{1}{12} \exp \left(\sum_{s=1}^t \Delta d_{s,j} + d_{0,j} \right) = g_{t-1,j} \exp \left(\bar{d}_j + \epsilon'_{t,j} \right), \quad t = 1, 2, \dots, T. \quad (43)$$

Data and estimation

Our data contains annual greenhouse gas intensity (in kilo tons CO₂eq per million \$) for each of the eleven sectors of the S&P 500 over the period 2014 to 2022 (9 data points). The greenhouse gas intensity is computed with (27) using greenhouse gas emissions and the market capitalization for each sector, obtained from Bloomberg. Greenhouse gas emissions are the sum of the first two scopes of greenhouse gas emissions (Scope 3 is omitted since data for such scope is less reliable). Table 2.4 presents descriptive statistics for yearly greenhouse gas intensities in each sector.

Table 2.4: Summary statistics for yearly greenhouse gas intensities of each sector of the S&P 500.

Sectors	Mean	StD	Skewness	Kurtosis	Min	Q1 25%	Q3 75%	Max
ENG	545.946	131.604	0.357	-0.096	351.005	464.142	606.477	780.901
HEC	6.520	2.562	0.464	-0.700	3.258	4.290	7.831	10.886
IND	164.350	82.198	0.157	-1.844	69.540	81.327	234.753	280.328
MAT	281.634	44.451	0.123	0.754	204.293	264.064	300.038	360.220
FIN	3.113	3.260	1.800	3.499	0.487	0.710	4.613	10.672
CST	40.573	12.254	0.079	-1.450	24.873	28.382	48.407	58.701
UTL	1303.112	566.755	0.666	0.090	636.523	794.194	1689.777	2383.883
ITC	8.137	3.601	-0.256	-1.477	2.742	5.139	11.420	12.760
RES	16.702	6.157	0.174	-0.074	7.135	12.438	19.441	27.374
CSV	16.682	9.899	0.234	-1.474	4.215	7.891	25.469	30.532
CDI	61.095	36.131	0.148	-1.202	15.623	29.517	84.635	116.708

The estimated parameters of the greenhouse gas intensity model (42) are reported in Table 2.5.

Table 2.5: Model parameter estimates for greenhouse gas intensities of sectors of the S&P 500.

Sector j	$\bar{c}_j$	σ_j^2
ENG	-0.068	0.078
HEC	-0.110	0.079
IND	-0.174	0.073
MAT	-0.032	0.027
FIN	-0.386	0.096
CST	-0.083	0.034
UTL	-0.165	0.028
ITC	-0.178	0.025
RES	-0.168	0.031
CSV	-0.248	0.114
CDI	-0.251	0.030

2.4.3 Trade-off between financial and environmental performance

Figure 2.1 shows the scatterplot of the historical Sharpe ratio and average log-GHG emission intensity for all sectors, which highlights the trade-off between GHG emissions and the risk-return profile offered by such sectors. A non-monotonic relationship is observed, with some sectors exhibiting both good (or both poor) Sharpe ratios and GHG intensities, such as Health Care (or Energy).

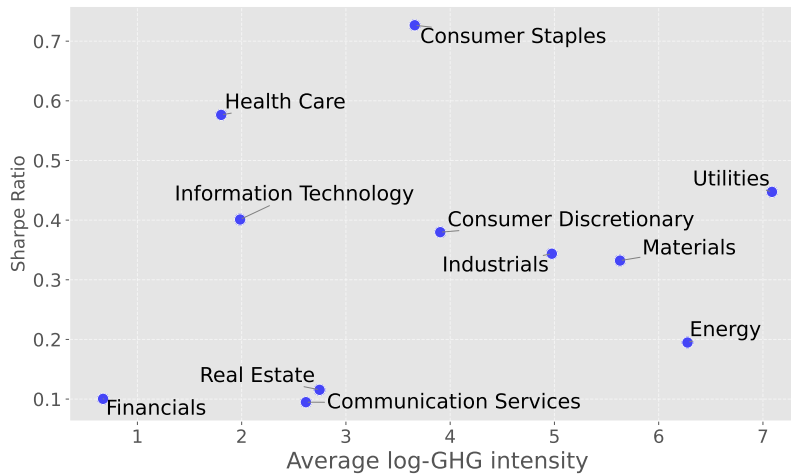


Figure 2.1: Historical value of annualized Sharpe ratio versus average log-GHG intensity.

2.5 Numerical Results

In this section, we apply the portfolio optimization model to determine the optimal dynamic asset allocation across the eleven S&P sectors in US stock market. The performance of optimal portfolios is then assessed throughout simulation experiments. The settings of the simulation are first provided. Subsequently, performance metrics for our emission-aware optimized portfolios are reported. Optimal allocations are then analyzed. Benchmarking against common investment strategies is also performed.

2.5.1 Portfolio optimization model setting

A 5-year time horizon ($T = 60$) is considered. Results are presented for 255 different investors who have various attitudes towards risk and GHG emissions, namely different parameters ρ (risk aversion) and γ (emissions aversion). More precisely, we consider 17 distinct values for ρ within the range $[0.005, 0.2]$ and 15 distinct values for γ within the range $[0, 0.25]$. These values are not uniformly spaced, as the training proved particularly sensitive when parameters are close to zero. Values considered are

$$\begin{aligned}\rho &\in \{0.005, 0.0075, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2\}, \\ \gamma &\in \{0.0, 0.0005, 0.0025, 0.005, 0.0075, 0.01, 0.02, 0.025, 0.04, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25\}.\end{aligned}$$

State variables $\mathbf{x}_t$ used to perform allocation decisions include: (i) the investment time t , (ii) the cumulative portfolio return W_t/W_0 , (iii) elements in the upper triangle of $\mathbf{H}_t$, which are conditional variances and covariances of the assets returns, (iv) the logarithm of the sectors' greenhouse gas intensities⁶ $\log(\mathbf{g}_t)$ and (v) the logarithm of previously emitted greenhouse gas over the portfolio wealth $\log(G_t/W_t)$. Elements in the upper triangle of $\mathbf{H}_t$ are standardized, which aids in improving the network's training efficiency.

⁶Here, the $\log(\cdot)$ of a vector is defined as the vector of the natural logarithm of each element.

The neural network dimension involves 80-dimensional⁷ inputs, one hidden layers with four neurons, and 11-dimensional outputs. Having only four neurons in the hidden layer is analogous to the use an auto-encoder, with the low capacity of the network reducing the risk of overfitting. This also provides the advantage of faster training. The training is performed with 25,000 simulated paths for sector returns and GHG intensities from the model outlined in Section 2.4. We use 5,000 epochs with early stopping, where training halts if the objective function shows no significant change after 200 epochs. Additionally, to improve learning efficiency, a scheduled learning rate is implemented. The ADAM algorithm from [74] is applied, with a learning rate parameter starting at 0.01 and halving as long as the objective function continues decreasing after sets of 5 epochs. The batch size for each iteration is set to $n_{\text{batch}} = 128$.

Whenever performance metrics are reported, such metrics are computed over another set of 25,000 path called the test set, as this allows to assess out-of-sample performance.

2.5.2 Performance analysis

The main quantity of interest to analyze performance over the entire time horizon is the portfolio's cumulative annualized return (CAR), expressed in percentage points. It is given by:

$$\text{CAR} = 100 \left[(W_T/W_0)^{\frac{12}{T}} - 1 \right].$$

The portfolio's financial performance is then assessed based on the following three metrics depicting its expected growth, its risk and its risk-reward tradeoff:

$$\text{Expected return} = \mathbb{E}[\text{CAR}],$$

$$\text{CAR standard deviation} = \text{Std}[\text{CAR}],$$

$$\text{Sharpe ratio} = \frac{\mathbb{E}[\text{CAR}] - r_f}{\text{Std}[\text{CAR}]},$$

⁷The breakdown of these 80 dimensions are: investment time (1-dimensional), cumulative portfolio return (1-dimensional), elements of the upper triangle of $\mathbf{H}_t$ (66-dimensional), sectors' GHG intensities (11-dimensional), and log wealth-normalized cumulative portfolio GHG emissions (1-dimensional).

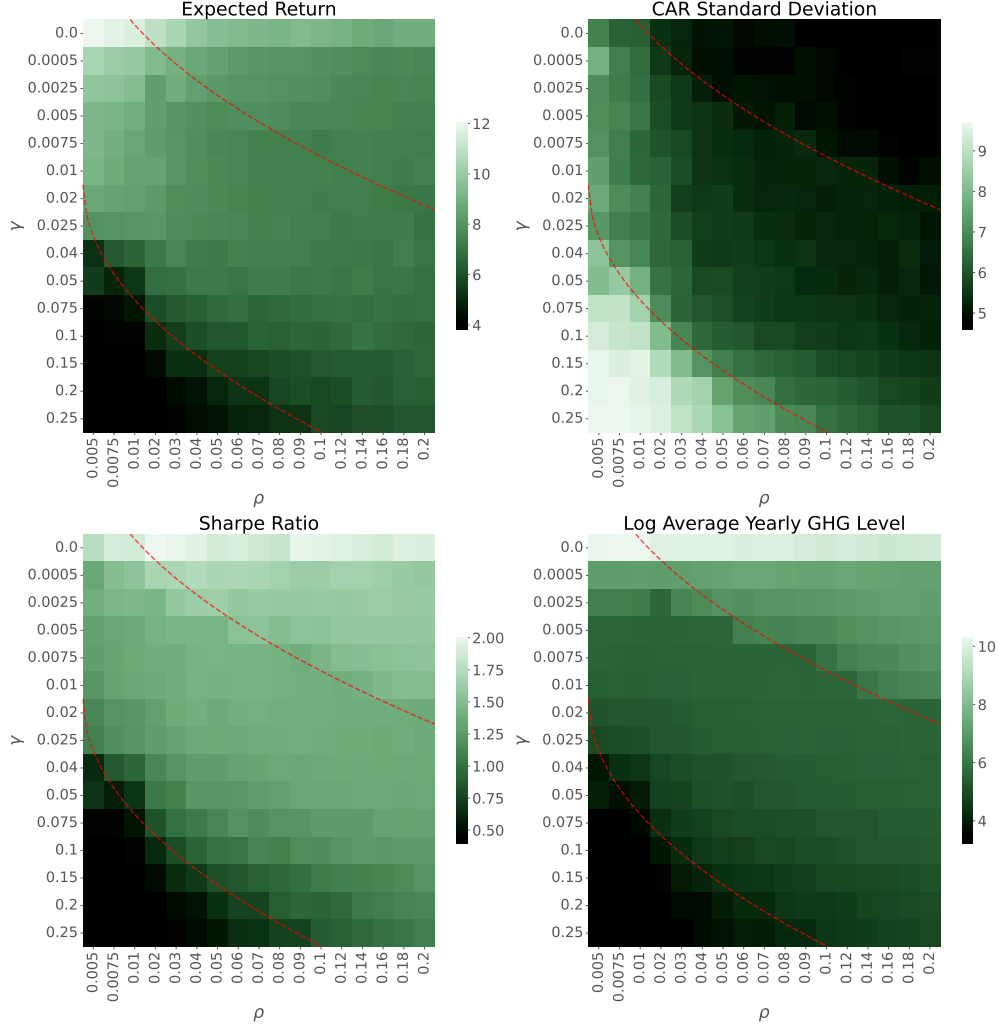


Figure 2.2: Heat maps of the expected value of the annualized portfolio return, its standard deviation, the associated Sharpe ratio, and the log average yearly GHG emissions of optimal portfolios for different risk and emission preference parameters ρ and γ .

where the risk-free rate is assumed to be $r_f = 0$ for simplicity. Furthermore, the environmental performance of the portfolio is represented by

$$\text{Log average yearly GHG emissions} = \log \left(\frac{G_T}{T/12} \right).$$

Heat maps from Figure 2.2 report, for the various investors, the above metrics for optimized portfolios. This provides insight into how changes in ρ and γ impact the trade-offs between financial performance and environmental considerations in the portfolio allocation process.

Based on the log-GHG heat map, we have segmented all four heat maps into three distinct regions using red dashed curves: portfolios with *Low GHG* (bottom-left corner), *High GHG* (upper-right corner) and *Mid GHG* (the rest) emissions.

The *Low GHG* portfolio region is marked by lower returns and higher risk (standard deviation), leading to a low Sharpe ratio. Investors in this region are socially responsible but less financially performant, prioritizing GHG emissions reduction. Although higher risk usually correlates with higher returns, the lower returns in this case stem from the inclusion of expected GHG levels in the objective function.

On the other hand, the *High GHG* region comprises portfolios of investors that are risk-averse, but comfortable with higher emission levels. These investors have higher Sharpe ratios, indicating a preference for assets with a favorable risk-return profile, despite their higher GHG emissions. This region's portfolios are mainly composed of sectors with higher Sharpe ratios.

The *Mid GHG* region represents investors who are willing to accept moderate financial performance alongside a moderate environmental impact. In this region, both the GHG emissions footprint and the Sharpe ratio are balanced, neither particularly high nor low. Investors from that region aim to optimize financial returns while managing both financial risk and environmental considerations. This strategy suits socially responsible investors who prioritize sustainability, but who are not prepared to fully sacrifice profitability to reduce GHG emissions. The *Mid GHG* region underscores the trade-offs between achieving financial goals and maintaining environmental responsibility.

Figure 2.3 represents the relationship between optimal portfolios' Sharpe ratio and log average GHG emissions for the various investors. There is a strong positive correlation (0.90) between the Sharpe ratio of optimal portfolios and their log average yearly GHG emissions.

Figure 2.4 displays (in two angles) the *optimal efficient surface*, which represents a set of investment portfolios yielding the highest expected returns for a given level of CAR standard deviation and log average GHG emissions. The efficient frontier surface slopes upward as log-GHG emissions increase, highlighting the cost of reducing GHG emissions in terms of expected return.

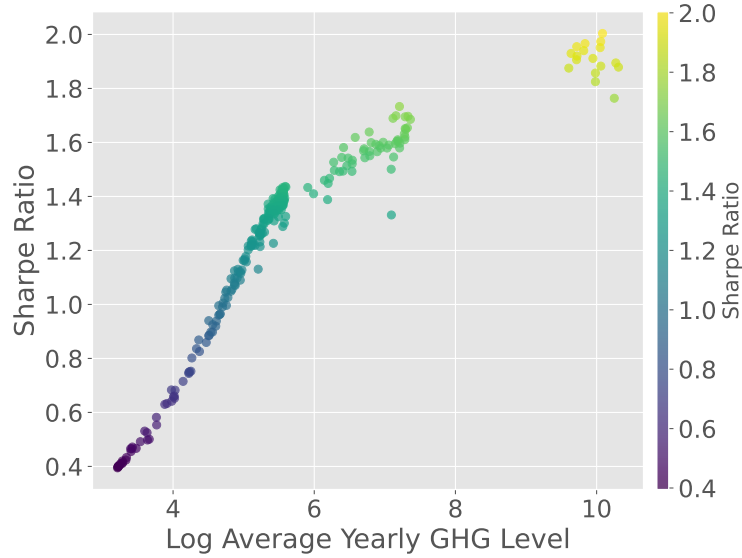


Figure 2.3: Sharpe ratio of optimal portfolios versus log-GHG emissions

Portfolios with low GHG emissions are associated with both lower expected returns and higher financial risk. The surface highlights the nonlinear trade-offs, where even small changes in GHG emissions can result in significant variations in risk and expected return. For instance, a slight rise in log-GHG emissions may cause a steep increase in expected returns, emphasizing the non-linear relationship between financial performance and environmental impact.

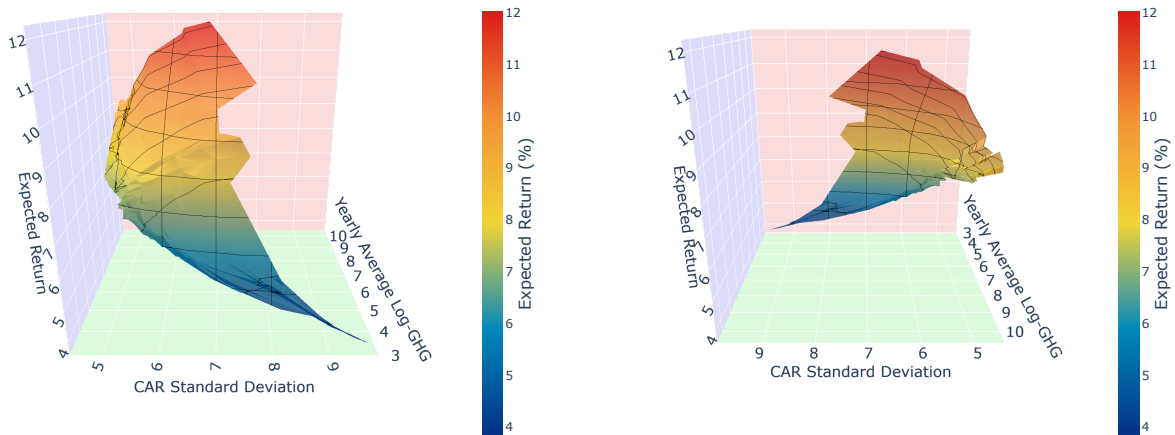


Figure 2.4: Optimal efficient surface of portfolios in two different views.

To quantify the trade-offs involved in setting up socially responsible investing strategies,

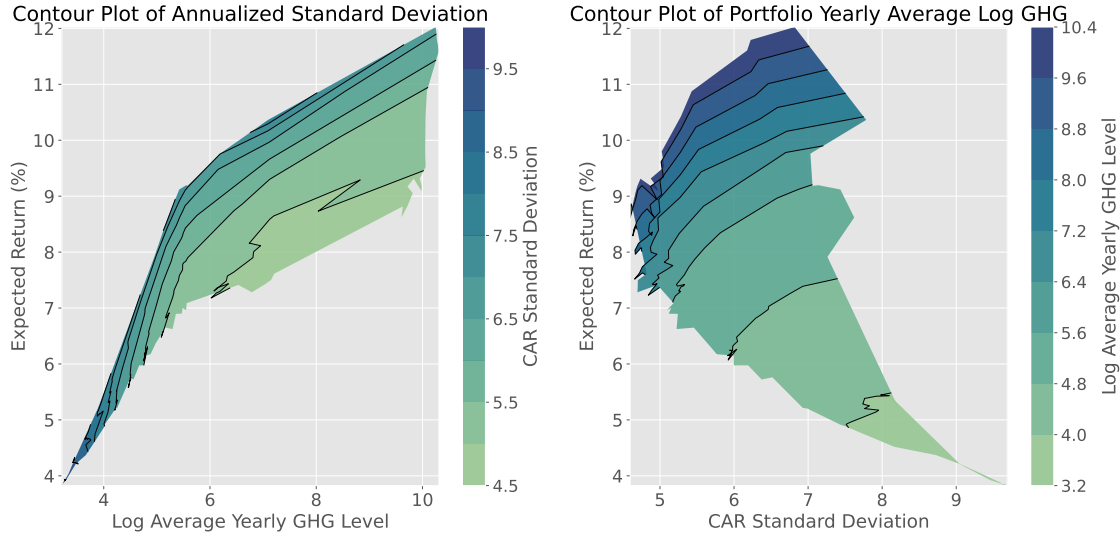


Figure 2.5: Contour plots illustrating the trade-offs between expected return, risk and emissions

Figure 2.5 presents contour plots of the expected return (i) versus log average GHG emissions for various levels of CAR standard deviation, or (ii) versus CAR standard deviation for different levels of log average GHG emissions. These plots basically present efficient surfaces of two of the variables when the third one is fixed, with each black curve representing a different value for the fixed variable. The black curves are not strictly monotonous due to the presence of training noise (i.e. slightly suboptimal training) for some of the agents.

The left panel in Figure 2.5 highlights the relationship between the expected return and log average yearly GHG emissions for different fixed levels CAR standard deviation. Portfolios with low emissions have both lower expected returns and higher risk. However, we observe that the slope of the expected return versus log-emissions decreases as emissions increase, indicating that the benefits of relaxing emission constraints are lower for high-emission investors than for low-emission investors. Indeed, for low-emission investors, small reductions in emissions lead to substantial increases in volatility and decreases in expected returns.

The right contour plot in Figure 2.5 illustrates the relationship between optimal the expected return and the CAR standard deviation for different emission levels. For fixed levels of GHG, the shape of the relationship between optimal expected return versus risk is a convex curve, which is reminiscent of Markowitz-type optimal portfolios. Thus, increasing expected returns becomes

increasingly costly in terms of additional risk as the portfolio risk (and expected return) increases.

2.5.3 Optimal allocation analysis

Figure 2.6 presents the heat maps of average portfolio weights allocated to different sectors based on varying levels of risk aversion (ρ) and emissions aversion (γ). Each cell in these heat maps represents the average weight assigned to a particular sector over the 5-year period and the various simulated paths for a given combination of ρ and γ .

For sectors such as Utilities, Energy, Industrials, and Materials, significant allocations are observed when investors disregard GHG emissions, i.e., when $\gamma = 0$. This is consistent with the data presented in Figure 2.1, which shows that these sectors are associated with high levels of GHG emissions, while they provide favorable risk-return trade-offs. Conversely, sectors Consumer Staples and Communication Services are also allocated to portfolios of investors having a moderate appetite for GHG emission reduction. Indeed, the Consumer Staples sector is characterized by medium GHG emissions and a high Sharpe ratio. This makes it an attractive choice for investors willing to accept some degree of environmental impact in exchange for potentially better risk-adjusted returns. Sectors Communication Services, Information Technology, Health Care, Real Estate, and Financials are widely distributed across different investor profiles, reflecting their ability to offer balanced combinations of return, risk, and GHG emissions under varying conditions. The Consumer Discretionary sector, however, is scarcely involved in any portfolio, indicating that it does not meet the risk-return or GHG preferences of most investor profiles considered in this analysis, based on our model.

One particularly interesting observation is the allocation of the Financials sector. Due to its low GHG emissions, it is primarily included in the *Low GHG* region of the heat map. However, it is also associated with a relatively low return-to-risk ratio, indicating that while it is a preferred choice for environmentally conscious investors, it may not be the optimal choice for those seeking higher returns.

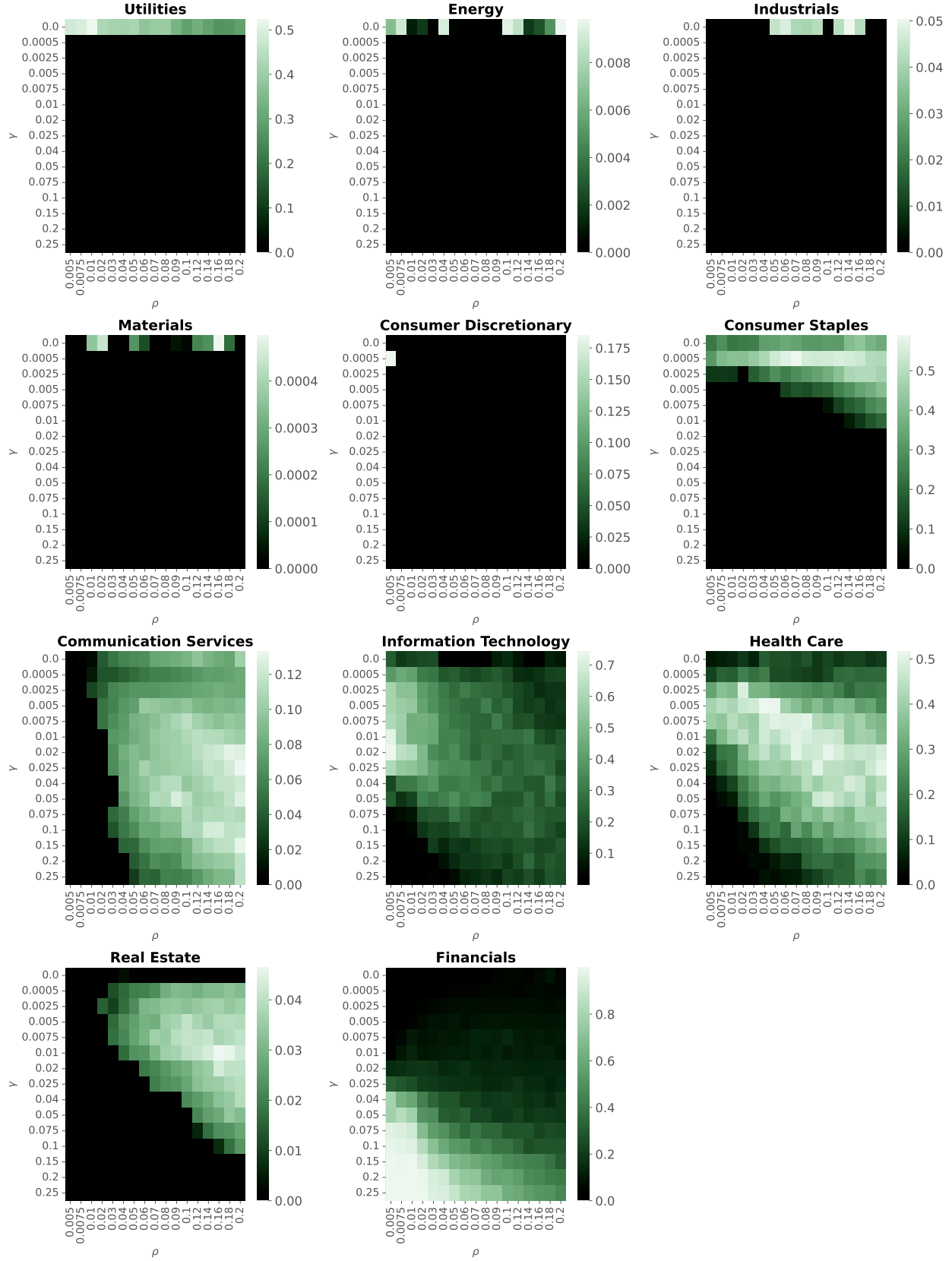


Figure 2.6: Heat maps of portfolio weights for different sectors

2.5.4 Performance Comparison with Benchmarks

We compare the performance of optimal portfolios (referred to as RL portfolios) with that of three benchmarks. The first benchmark is an equal-weighted portfolio, where the weights of all eleven sectors are set to $\frac{1}{11}$. The second benchmark is a static mean-variance-GHG optimization. In this context, the objective function of the optimization problem still is (30), but the portfolio weights are fixed over the entire time horizon, without dynamic adjustments over time. Weights are therefore optimized in a classic fashion without relying on reinforcement learning. The third benchmark is a myopic dynamic mean-var-GHG optimization. In this scheme, weights are dynamically adjusted by re-solving a one-period ahead optimization problem on each period. This approach is explained in Section 2.6.

The various performance metrics are computed for the benchmark strategies and compared to these of optimal portfolio for three investors:

- Investor 1: has high tolerance for both risk and GHG emissions ($\rho = 0.005$ and $\gamma = 0.0$),
- Investor 2: has high aversion for both risk and GHG emissions ($\rho = 0.2$ and $\gamma = 0.2$),
- Investor 3: has moderate aversion for risk and GHG emissions ($\rho = 0.07$ and $\gamma = 0.01$).

Figure 2.7 compares objective function components (expected wealth, its variance, and cumulative GHG emissions) for the optimal strategy to these of benchmarks. Each row of Figure 2.7 represents one of the three investors, whereas each columns depicts a different objective function component. The components are reported for all time horizons up until $T = 60$, despite the agents being trained with a $T = 60$ horizon.

The first row of Figure 2.7 illustrates that the RL portfolio of the Investor 1, which is highly tolerant to risk and emission levels, significantly outperforms all three benchmarks in terms of wealth accumulation. Conversely, the optimal portfolio produces higher variance and GHG emissions than the benchmarks, which is consistent with the fact that risk and GHG emissions are not material concerns for Investor 1. In this case, the optimal portfolio proves to be the most effective choice for an investor focused solely on maximizing returns without considering risk or environmental impact.

For Investor 2 having high aversion for both risk and emissions, the optimal RL portfolio dominates the static portfolio, with all three components (wealth, variance and emissions) being more favorable for the former. Whereas the expected wealth is marginally lower for the RL portfolio than for the equal-weighted benchmark, both its variance and GHG emissions are substantially lower. Although GHG emissions of the myopic dynamic benchmark are slightly lower than these of the RL portfolio, the RL approach dominates the myopic dynamic benchmark both in terms of expected accumulated wealth and wealth variance.

Finally, concerning Investor 3 who exhibits moderate aversion to risk and emissions, the RL portfolio outperforms all three benchmarks in terms of expected return and risk, and even produces lower emissions than the equal-weighted benchmark. The latter strategy is therefore dominated by the optimal strategy in that case. The optimal portfolio is therefore able to attain a sound balance between financial and environmental performance.

When analyzing the evolution of performance over the various time horizons, the RL portfolio stands out for its stability; it consistently performs well throughout the entire investment period, not just at the final time horizon, despite the objective function only focusing on terminal performance. Higher performance at the end of the investment horizon does not come at the expense of good performance at intermediary time steps, which is valuable for investors who seek reliable returns and controlled volatility over time.

Figure 2.8 shows the distribution of the expected return, the CAR standard deviation, and log average yearly GHG emissions over all investors considered in this study (i.e. for all combinations of values for ρ and γ outlined in Section 2.5.1). Distributions for RL portfolios are compared to these of benchmarks. Because the equal-weighted strategy is identical for all investors, such benchmark produces a degenerate distribution with a single point. The plots highlight the clear out-performance of the optimal RL portfolios over the static, myopic dynamic and equal-weighted benchmarks. The RL strategy delivers substantially higher expected annualized returns while maintaining lower volatility than the static approach for comparable levels of emissions. Moreover, GHG emissions are much lower for RL portfolios than for the equal-weighted one. The RL approach also offers

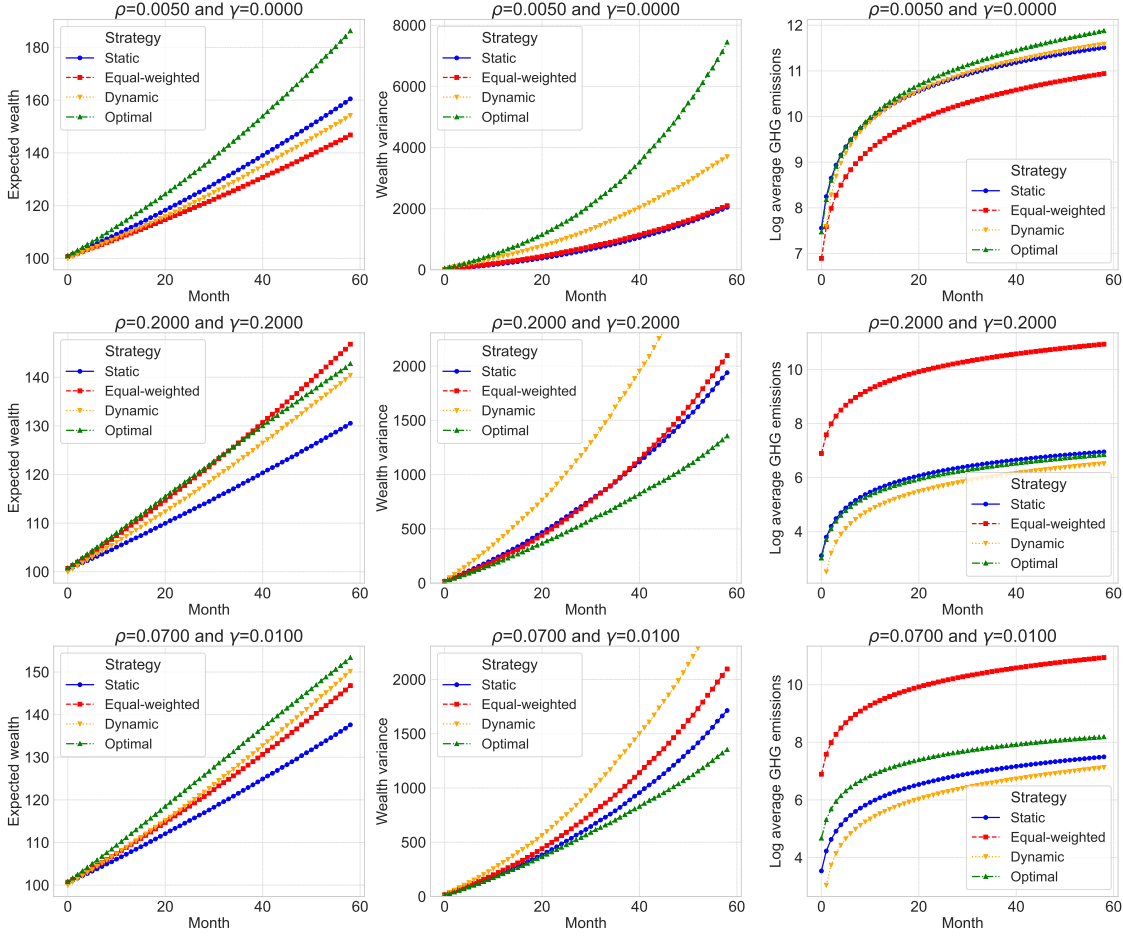


Figure 2.7: Comparison of expected wealth, wealth variance and emissions for the RL (Optimal) the equal-weighted, the optimal static (Static), and the myopic dynamic (Dynamic) portfolios for three different investors)

substantial risk reduction in comparison to myopic dynamic portfolios. The RL approach is thus the most effective choice for investors with diverse preferences related to financial performance and sustainability.

2.5.5 Regression Analysis

In this section, analyze how portfolio weights adjust in response to changes in market conditions. We first define market volatility as the average volatility over all sectors: $V_{m,t} = \frac{\sum_{i=1}^{11} \sqrt{h_{t,i}}}{11}$. This results in 25,000 different paths for market volatility. Next, we perform the following regression

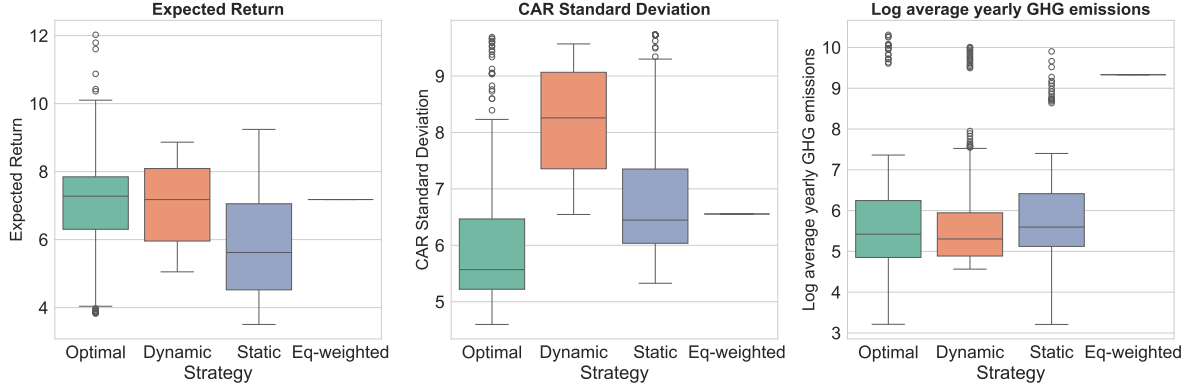


Figure 2.8: The distribution of expected return, CAR standard deviation, and log average yearly GHG emissions over the different investors (the different combinations of values for ρ and γ outlined in Section 2.5.1) for the following four strategies: RL (Optimal), optimal static (Static), myopic dynamic (Dynamic), and Equal-weighted.

across all paths: for all sectors $i = 1, \dots, 11$,

$$\omega_{i,t}^{(n)} = \alpha_i + \beta_i V_{m,t}^{(n)} + \xi_{i,t}^{(n)} \quad n = 1, \dots, 25,000, \quad t = 1, \dots, T, \quad (44)$$

where $\omega_{i,t}^{(n)}$ represents the weight of sector i at time t for path n , and $\xi_{i,t}^{(n)}$ is the corresponding noise term.⁸ Table 2.6 shows the slope and the associated t-statistics for each sector's regression, with the Investor 3 (moderate aversion for financial risk and emissions) being considered to determine portfolio weights. When market volatility increases in a given month, the weights allocated to the Communication Services and Health Care increase, while the weights for Information Technology, Real Estate, and Financials decrease. As shown in Table 2.1, the standard deviation of returns of the former sectors is lower than that of the latter sectors. During volatile periods, the optimal strategy thus consists in reducing investments in sectors exhibiting higher volatility. This highlights that during a crisis or a period of financial turbulence, the optimal investor acts defensively and temporarily increases its exposure to assets associated with higher emissions. A return to higher exposure to more environmentally-friendly assets would occur when normal economic conditions resume. This behavior differentiates the RL portfolio from the static benchmark, as static strategies

⁸The regression is performed out-of-sample on paths that have not been used for the training of the neural network recommending allocation weights.

do not adjust to changes in market volatility.

Sectors	Slope	t-statistics
Utilities	7×10^{-6}	6.8
Energy	5×10^{-6}	7.0
Industrials	4×10^{-6}	7.4
Materials	8×10^{-6}	7.0
Consumer Discretionary	1.7×10^{-5}	4.3
Consumer Staples	2.2×10^{-5}	3.9
Communication Services	5.44	147.0
Information Technology	-4.57	-99.8
Health Care	10.43	191.1
Real Estate	-0.43	-18.0
Financials	-10.87	-249.2

Table 2.6: Slopes and associated t-statistics in the regression (44) for each sector, which regresses portfolio weights of Investor 3 (moderate aversion for financial risk and emissions) on market volatility. An intercept has been included in each regression, but it is not reported.

2.6 Dynamic Portfolio Optimization

This section presents the construction approach for myopic dynamic portfolio. At each time t , we find weights ω_t which maximize

$$\mathbb{E}_t[W_{t+1}] - \rho \text{Var}_t(W_{t+1}) - \gamma \mathbb{E}_t[G_{t+1}]. \quad (45)$$

The various term in the above equation can be expressed explicitly as shown in what follows. First,

$$\mathbb{E}_t[W_{t+1}] = \mathbb{E}_t[W_t \omega_t^\top (\mathbf{r}_{t+1} + 1)] = W_t \omega_t^\top \mathbb{E}_t[(\mathbf{r}_{t+1} + 1)].$$

From (36), $\log(\mathbf{r}_{t+1} + 1)$ has a multivariate normal distribution, and thus $(\mathbf{r}_{t+1} + 1)$ has a multivariate lognormal distribution. Therefore,

$$\mu_t^l = \mathbb{E}_t[(\mathbf{r}_{t+1} + 1)] = \left(\exp(\mu_{t+1,1} + 0.5H_{t+1,11}), \dots, \exp(\mu_{t+1,N} + 0.5H_{t+1,NN}) \right)^\top.$$

The variance can be expressed as

$$\text{Var}_t(W_{t+1}) = \text{Var}_t(W_t \boldsymbol{\omega}_t^\top (\mathbf{r}_{t+1} + 1)) = W_t^2 \boldsymbol{\omega}_t^\top C_t \boldsymbol{\omega}_t$$

where $C_t = \text{Var}_t(\mathbf{r}_{t+1} + 1)$ and for any i, j :

$$C_{t,ij} = \exp\left(\mu_{t+1,i} + \mu_{t+1,j} + 0.5(H_{t+1,ii} + H_{t+1,jj})\right)(e^{H_{t+1,ij}} - 1)$$

Expected emissions can be expressed as

$$\mathbb{E}_t[G_{t+1}] = G_t + \boldsymbol{\omega}_t^\top \mathbb{E}_t[g_{t+1}] W_t = G_t + \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_t^g W_t. \quad (46)$$

According to (43), for each j ,

$$\mu_{t,j}^g = \mathbb{E}_t[g_{t+1,j}] = \mathbb{E}_t\left[g_{t,j} \exp\left(\bar{d}_j + \epsilon'_{t+1,j}\right)\right] = g_{t,j} \exp\left(\bar{d}_j + 0.5 \frac{\sigma_{\epsilon,j}^2}{12}\right).$$

Now we can rewrite the objective function (45) as

$$\mathbb{E}_t[W_{t+1}] - \rho \text{Var}_t(W_{t+1}) - \gamma \mathbb{E}_t[G_{t+1}] = W_t \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_t^l - \rho W_t^2 \boldsymbol{\omega}_t^\top C_t \boldsymbol{\omega}_t - \gamma(G_t + \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_t^g W_t).$$

Removing γG_t from the above since it is fixed, and dividing what remains by W_t , the portfolio optimization problem reduces to

$$\begin{aligned} \max_{\boldsymbol{\omega}_t} \quad & \boldsymbol{\omega}_t^\top (\boldsymbol{\mu}_t^l - \gamma \boldsymbol{\mu}_t^g) - \rho W_t \boldsymbol{\omega}_t^\top C_t \boldsymbol{\omega}_t \\ \text{s.t.} \quad & \mathbf{1}^\top \boldsymbol{\omega}_t = 1, \\ & \omega_{t,j} \geq 0. \end{aligned} \quad (47)$$

This is a constrained quadratic problem which we solve through numerical optimization.

Chapter 3

Learning to Hedge Swaptions

3.1 Introduction

Interest rate derivatives play a central role in modern financial markets, with interest rate swaps and swaptions serving as foundational instruments for risk management and asset-liability transformation. Swaptions—options on interest rate swaps—are particularly vital for hedging interest rate volatility and convexity risk, offering flexibility in managing contingent liabilities and exposures. They are widely used by institutions such as insurance companies, pension funds, and structured product issuers, and are integral to the pricing and risk management of callable bonds, mortgage-backed securities, and other complex instruments. According to the Bank for International Settlements [18], the global notional value of outstanding interest rate derivatives exceeds \$540 trillion, with swaptions comprising over \$46 trillion of this total, reflecting their deep integration into the financial ecosystem.

Multiple models have been developed to describe the dynamics of the interest rate market. The fundamental goal of these models is to capture realistic market behavior, ensuring that the pricing and hedging of interest rate instruments closely aligns with observed prices. Broadly, these models can be categorized into two principal classes: the short (spot) rate class and the forward rate class. Notable examples of short rate models include those by Ho and Lee [61], Hull and White [64],

and Vasicek [108]. Prominent representatives of forward rate models are the Heath-Jarrow-Morton (HJM) framework, the lognormal LIBOR market model (LMM), and the swap market model. Each class has evolved from basic affine specifications to more sophisticated structures, introducing features such as time-inhomogeneous volatility and multi-factor extensions to better capture realistic interest rate dynamics. These enhancements are particularly important for hedging, as they enable more accurate representation of the myriad risk factors that affect the pricing and effective risk management of swaptions and other interest rate derivatives.

Traditional hedging approaches in interest rate derivatives rely heavily on factor-based term structure models such as the Hull and White [64] approach, Gaussian multi-factor models [24], and Libor Market Model (LMM) frameworks [22]. These models serve as the foundation for the construction of hedging strategies that aim to neutralize first-order sensitivities, and in some cases, volatility risk.

Under the Hull–White one-factor model, hedging is typically based on duration matching, which provides analytical tractability but suffers from significant limitations in capturing non-parallel yield curve shifts and volatility. Multi-factor Gaussian models improve upon this by introducing level, slope, and curvature factors, thereby reducing residual risk. Market-consistent models such as the LMM offer superior performance in replicating market-observed volatilities and facilitate delta-vega hedging, yet they require frequent recalibration to a dense swaption volatility surface, making them computationally expensive and prone to instability [91].

Despite these advancements, conventional hedging strategies—whether based on fixed ratios or dynamically updated sensitivities—remain fundamentally local in nature. They typically optimize hedge positions one period at a time, without accounting for compounding risk premia or path-dependent exposures. Deep reinforcement learning (DRL), as in Buehler et al. [25], offers an alternative to such conventional schemes. In our framework, DRL leverages parametric yield-curve dynamics while optimizing policies jointly over multiple periods. While DRL training environments are typically built on model-based simulations, the agent itself does not impose parametric restrictions on hedge ratios. Instead, it learns flexible, state-dependent hedging policies

through interaction with simulated or historical market environments. The advantages of DRL-based hedging lie in enabling multi-stage optimal control, flexibility in targeting different objective functions, and the ability to account for physical distributional features such as risk premia. By dynamically adapting its hedge ratios over time, DRL effectively captures higher-order and path-dependent risk exposures that traditional approaches fail to address.

Literature reveals that single-factor models are often insufficient for hedging interest rate derivatives, whereas two- or three-factor models significantly enhance hedging efficiency by more accurately capturing yield curve movements and volatility dynamics [39]. To the best of the authors' knowledge, with the exception of [94], there are currently no other published works that employ deep reinforcement learning for hedging interest rate derivatives. This study aims to contribute to this gap by developing a DRL-based framework for hedging European swaptions when the spot rate dynamic follows a multi-factor short rate process, specifically the discrete-time arbitrage-free Nelson–Siegel model of Eghbalzadeh et al. [42], which is a discrete-time version of the arbitrage-free Nelson–Siegel model of Christensen et al. [30]. Such a paradigm shift represents an important advancement in tackling the unique challenges posed by interest rate derivatives hedging and in broadening the scope of neural-network-based hedging strategies in this domain.

This paper makes several contributions to the literature on dynamic hedging of interest rate derivatives. First, we develop a DRL framework for swaption hedging that departs from traditional sensitivity-based strategies. By formulating hedging as a sequential decision problem, our approach allows agents to internalize pathwise dynamics and transaction costs, resulting in more flexible and adaptive strategies. Second, we demonstrate how different risk preferences can be encoded through alternative objective functions. In particular, we train agents under mean squared error, downside risk, and Conditional Value-at-Risk (CVaR), and show how each criterion induces distinct hedging styles. This comparative analysis provides new insights into the trade-offs between hedge effectiveness, tail-risk protection, and premium harvesting. Third, we evaluate the marginal value of multiple hedging instruments and identify distinct roles for the underlying and complementary

swaps. Our results show that most hedging effectiveness is achieved with two swaps, with the underlying swap serving as the primary replication instrument and the complementary swap absorbing residual yield curve exposures or acting as a tail-risk hedge. Finally, we highlight the structural advantages of RL-based hedging over rho-hedging, including lower transaction costs, robustness to model misspecification, and general superior performance. These findings underscore the potential of reinforcement learning as a robust and economically meaningful framework for interest rate derivatives risk management.

The remainder of this paper is organized as follows. Section 3.2 reviews the relevant literature on swaption hedging and recent advances in DRL for hedging. Section 3.3 describes the market environment, the DTAFNS model, and the associated pricing of interest rate instruments such as swaps and swaptions. Section 3.4 introduces the proposed hedging framework, including the construction of the replicating portfolio, the deep hedging approach, and benchmark strategies based on factor sensitivities. Section 3.5 details the experimental design, and neural network architectures employed for both pricing and hedging tasks. Section 3.6 presents the empirical results, comparing the performance of DRL-based hedging strategies against traditional benchmarks under multiple risk metrics. Section ?? concludes.

3.2 Literature Review

Hedging swaptions presents enduring challenges due to the complex nature of interest rate dynamics and modeling choices. The literature can be structured into three key categories: classical methods, advances in term structure modeling, and recent innovations driven by machine learning.

Jamshidian [68] demonstrates that a swaption can be replicated using a finite number of zero-coupon bonds within the log-normal LIBOR market model (LMM), establishing a direct connection between swaptions and fundamental fixed-income contracts. Brace et al. [23] compute the Greeks of swaptions under this model and show that the resulting hedging strategy aligns closely with the Black swaption formula, provided the covariance matrix of the model is of rank one. The

effectiveness of delta hedging strategies within the LMM is extensively explored in Tim Dun et al. [103], who employ Monte Carlo simulations to show that while at-the-money swaptions can be hedged effectively, significant errors persist for in- and out-of-the-money contracts due to model limitations and sensitivity to volatility and payoff asymmetry.

Research on term structure models highlights the advantages of multi-factor frameworks. Driessen et al. [39] finds that multi-factor term structure models outperform single-factor models for hedging interest rate derivatives such as caps and swaptions because they more accurately capture yield curve movements. Fan et al. [45] demonstrate that higher-order multifactor HJM models significantly reduce hedging errors in the swaption market, while Fan et al. [46] further show that at least two factors are required to robustly capture the yield curve dynamics necessary for effective swaption hedging. An and Suo [6] compare one-factor LIBOR and Swap Market Models, concluding that although they yield comparable results, significant biases and limitations underscore the need for more sophisticated or multi-factor methods. In addition, Cresnik [33] evaluates the hedging performance of the Black [19], Hull and White [64], and SABR [57] models using both historical and synthetically generated scenarios—finding that stochastic volatility and multifactor models such as SABR provide superior risk reduction versus simpler one-factor alternatives.

Recent research applies machine learning to address the complexity and high dimensionality of swaption hedging. Wang et al. [111] introduce a deep learning-based BSDE solver within the LMM, utilizing a backward neural network architecture that efficiently prices and hedges Bermudan swaptions, computes Greeks, and handles high-dimensional problems. More recently, Oya [94] proposes a deep hedging framework for Bermudan swaptions. They rely on the CVaR convex risk measure, achieving improved downside risk control over traditional arbitrage-free methods and increasing robustness in scenarios with realistic market frictions.

Multi-factor short rate models provide a richer representation of yield curve dynamics, allowing different maturities to move independently and thus capturing the complex shifts, twists, and curvature observed in real markets [39, 56]. Empirical evidence shows that these models achieve superior hedging performance for swaptions compared to one-factor models, notably reducing

hedging errors when the number of hedge instruments matches the number of model factors [39, 45].

Within the universe of multi-factor term structure models, the Nelson-Siegel framework stands out as the most widely adopted and practically relevant approach in applied finance. This model is highly valued in practice for its ease of estimation, intuitive interpretation, and widespread use by central banks and financial institutions to fit and forecast yield curves across a range of maturities. Originally introduced by Nelson and Siegel [92] as a parsimonious parametric method for fitting the yield curve, the Nelson-Siegel model remains popular due to its simplicity and effectiveness in capturing level, slope, and curvature using only a handful of parameters. The extension by Diebold and Li [35] to a dynamic setting allowed these three factors to evolve over time, enabling the model to serve for forecasting and risk management, but the initial dynamic approach did not satisfy the no-arbitrage condition fundamental to financial theory. This gap was filled by Christensen et al. [30], who formulated the arbitrage-free Nelson-Siegel (AFNS) model and thus combined empirical tractability with theoretical rigor. More recently, Eghbalzadeh et al. [42] introduced the discrete-time arbitrage-free Nelson-Siegel (DTAFNS) model, classified by Nawalkha et al. [91] as an $A_0(3)$ term structure model. The DTAFNS model further addresses computational challenges by making simulation and estimation more tractable and aligns well with the fact that most financial data are observed at discrete intervals.

Deep hedging, powered by deep reinforcement learning (DRL), has emerged as an advanced framework for managing the risks of financial derivatives in environments with transaction costs, stochastic volatility, and other real-world market frictions. Unlike traditional model-based hedging methods, such as Black-Scholes delta hedging, deep hedging employs data-driven neural networks that sequentially optimize hedging actions using observed market states. The foundational contribution by Buehler et al. [25] demonstrates the capacity of DRL architectures to efficiently hedge portfolios under realistic constraints, while subsequent studies such as Cao et al. [26] successfully extend these frameworks to incorporate transaction costs, consistently outperforming classical strategies in synthetic environments.

Further empirical research [89, 54] confirmed these findings on real market data, underscoring

the robust transferability of DRL-trained agents to novel market regimes and the effectiveness of neural-policy hedgers in reducing overall hedging costs. For a comprehensive overview of methodologies and trends in this field, see the survey by Pickard and Lawryshyn [97], which analyzes DRL models, state/action spaces, and reward structures across 17 major studies.

3.3 Market Environment and Financial Instruments

We assume that the short-rate dynamics follow the DTAFNS model proposed by Eghbalzadeh et al. [42]. The DTAFNS model offers a tractable, arbitrage-free framework for yield curve modeling in discrete time. Its closed-form solutions, empirical robustness, and compatibility with financial applications make it a valuable tool for pricing, risk management, and forecasting in stochastic interest rate environments.

The market operates in discrete time with time steps $t = 0, 1, \dots, T$, each separated by a fixed time increment Δ (in years). The financial market is modeled on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ denotes the real-world (physical) probability measure and $\mathcal{F} := \{\mathcal{F}_t\}_{t=0}^T$ represents the natural filtration encoding the information flow in the market. The time- t term structure is driven by three factors, $X_t^{(1)}$, $X_t^{(2)}$ and $X_t^{(3)}$, each of which is $\mathcal{F}_t$ -measurable. The short rate r_t prevailing over the interval $[t, t+1)$ is specified as

$$r_t = X_t^{(1)} + X_t^{(2)}. \quad (48)$$

The factors vector $X_t = [X_t^{(1)}, X_t^{(2)}, X_t^{(3)}]^\top$ is latent and evolves under measure $\mathbb{P}$ according to

$$X_{t+1} = X_t + \kappa^\mathbb{P}(\theta^\mathbb{P} - X_t) + \Sigma Z_{t+1}^\mathbb{P}. \quad (49)$$

Here, $\kappa^\mathbb{P} = [\kappa_{i,j}^\mathbb{P}] \in \mathbb{R}^{3 \times 3}$ is the mean-reversion parameters matrix, $\theta^\mathbb{P} = [\theta_1^\mathbb{P}, \theta_2^\mathbb{P}, \theta_3^\mathbb{P}]^\top$ is the long-term level of the factors, and $\Sigma = [\Sigma_{i,j}] \in \mathbb{R}^{3 \times 3}$ is a diagonal volatility matrix with strictly positive entries. The innovations $\{Z_t^\mathbb{P}\}_{t=1}^T$ are i.i.d. standard multivariate Gaussian vectors under $\mathbb{P}$ with

contemporaneous correlations

$$\rho_{ij}^{(Z)} := \text{corr}(Z_{t,i}^{\mathbb{P}}, Z_{t,j}^{\mathbb{P}}), \quad \text{for } i, j = 1, 2, 3.$$

The initial factor value $X_0 = x_0 \in \mathbb{R}^3$ is assumed known and fixed.

To enable arbitrage-free pricing of interest rate derivatives such as bonds, bond options, swaps, and swaptions, Eghbalzadeh et al. [42] specify risk-neutral dynamics under an equivalent martingale measure $\mathbb{Q} \sim \mathbb{P}$. The factor dynamics under $\mathbb{Q}$ is assumed to be of the form

$$X_{t+1} = X_t + \kappa^{\mathbb{Q}}(\theta^{\mathbb{Q}} - X_t) + \Sigma Z_{t+1}^{\mathbb{Q}}, \quad (50)$$

where $\kappa^{\mathbb{Q}}$ and $\theta^{\mathbb{Q}}$ are obtained by solving the following system

$$\begin{cases} \kappa^{\mathbb{Q}} = \kappa^{\mathbb{P}} - \Sigma\gamma, \\ \kappa^{\mathbb{Q}}\theta^{\mathbb{Q}} = \kappa^{\mathbb{P}}\theta^{\mathbb{P}}, \end{cases} \quad (51)$$

for some constant vector γ , with $\{Z_t^{\mathbb{Q}}\}$ being independent standard Gaussian vectors with correlation matrix $[\rho_{ij}^{(Z)}]_{i,j=1}^3$ under $\mathbb{Q}$. This dynamics retains the same autoregressive structure as under the physical measure. This consistency is desirable as it facilitates analytical tractability and simulation under both real-world and risk-neutral measures. Furthermore, matrix $\kappa^{\mathbb{Q}}$ is assumed to be of the form

$$\kappa^{\mathbb{Q}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{bmatrix}, \quad \text{with } \lambda \in (0, 1).$$

The null first row of $\kappa^{\mathbb{Q}}$ entails that the first factor is non-stationary under $\mathbb{Q}$ despite being stationary under $\mathbb{P}$.

Zero Coupon Pricing. The DTAFNS model admits an exponential-affine solution for zero-coupon bond prices. Specifically, for a bond maturing at time T , the arbitrage-free price at time $t < T$ is

given by

$$P(t, T) = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(-\Delta \sum_{j=t}^{T-1} r_j \right) \middle| \mathcal{F}_t \right] = A_\tau \exp \left[-\Delta B_\tau^\top X_t \right], \quad (52)$$

where $\tau = T - t$ denotes the remaining time to maturity in discrete steps, and $B_\tau = [B_\tau^{(1)}, B_\tau^{(2)}, B_\tau^{(3)}]^\top$ is a vector of exponentially decaying loadings associated with each factor. The coefficient A_τ is a scalar capturing the convexity adjustment. Closed-form formulas for (A_τ, B_τ) are provided in Section 3.7.

Interest Rate Swaps Pricing. According to Brigo et al. [24], a Payer Forward-Starting Interest Rate Swap (PFS) is a contract with a future effective date T_α , where one party agrees to pay a fixed interest rate K on a nominal amount N and receive a floating rate over a sequence of dates $\{T_{\alpha+1}, \dots, T_\beta\}$, where $T_{i+1} = T_i + \Delta$ for all i . At each payment date T_i in this schedule, the fixed leg pays

$$NK\Delta,$$

while the floating leg pays

$$NL(T_{i-1}, T_i)\Delta,$$

where $L(T_{i-1}, T_i)$ denotes the simply-compounded spot rate set at T_{i-1} for the period $[T_{i-1}, T_i]$:

$$L(T_{i-1}, T_i) = \frac{1 - P(T_{i-1}, T_i)}{P(T_{i-1}, T_i)\Delta}.$$

The floating leg resets at dates $T_\alpha, T_{\alpha+1}, \dots, T_{\beta-1}$ and pays at dates $T_{\alpha+1}, \dots, T_\beta$. Let $\mathcal{T} := \{T_\alpha, \dots, T_\beta\}$ denote the full tenor schedule of the swap.

The discounted value of the PFS cash flows at a time $t < T_\alpha$ is given by

$$\sum_{i=\alpha+1}^{\beta} D(t, T_i) N \Delta (L(T_{i-1}, T_i) - K),$$

where $D(t, T_i) = \exp \left(-\Delta \sum_{s=t}^{T_i-1} r_s \right)$ is the stochastic discount factor between time t and the

payment date T_i . Using no-arbitrage arguments and risk-neutral pricing, the forward-start swap value can be expressed in terms of observable market instrument prices as

$$\begin{aligned} \text{PFS}(t, \mathcal{T}, N, K) &= N\Delta \sum_{i=\alpha+1}^{\beta} P(t, T_i) (F(t; T_{i-1}, T_i) - K) \\ &= N (P(t, T_{\alpha}) - P(t, T_{\beta})) - NK\Delta \sum_{i=\alpha+1}^{\beta} P(t, T_i), \end{aligned} \quad (53)$$

where $F(t; T_{i-1}, T_i)$ denotes the simply-compounded forward rate for the period $[T_{i-1}, T_i]$

$$F(t; T_{i-1}, T_i) = \frac{1}{\Delta} \left(\frac{P(t, T_{i-1})}{P(t, T_i)} - 1 \right).$$

The price of a Receiver Forward-Start Interest Rate Swap (RFS) is simply the negative of the PFS price. The forward swap rate, denoted $S_{\alpha, \beta}(t)$, is the fixed rate that equates the value of the PFS to zero (i.e., the par rate). Solving for this value yields

$$S_{\alpha, \beta}(t) = \frac{P(t, T_{\alpha}) - P(t, T_{\beta})}{\Delta \sum_{i=\alpha+1}^{\beta} P(t, T_i)}.$$

This rate is commonly used as a benchmark in swaption pricing and hedging applications.

Swaption Pricing. A *European payer swaption* is a financial derivative that grants its holder the right, but not the obligation, to enter into a payer interest rate swap (IRS) at a predetermined future time—typically coinciding with the first reset date of the underlying swap, denoted T_{α} . The length of the underlying swap, given by $T_{\beta} - T_{\alpha}$, is referred to as the *tenor* of the swaption.

From (53), the value of the underlying payer IRS at the swaption maturity T_{α} is

$$N\Delta \sum_{i=\alpha+1}^{\beta} P(T_{\alpha}, T_i) (F(T_{\alpha}; T_{i-1}, T_i) - K).$$

The swaption is exercised only if this value is positive. Thus, the discounted payoff of the payer

swaption at a time $t < T_\alpha$ is

$$ND(t, T_\alpha) \Delta \left(\sum_{i=\alpha+1}^{\beta} P(T_\alpha, T_i) (F(T_\alpha; T_{i-1}, T_i) - K) \right)^+.$$

The swaption is said to be *at-the-money* (ATM) if the strike equals the forward swap rate, i.e.

$$K = K_{\text{ATM}} := S_{\alpha, \beta}(0) = \frac{P(0, T_\alpha) - P(0, T_\beta)}{\Delta \sum_{i=\alpha+1}^{\beta} P(0, T_i)}.$$

It is *in-the-money* (ITM) for the payer swaption if $K < K_{\text{ATM}}$ and *out-of-the-money* (OTM) if $K > K_{\text{ATM}}$. The reverse applies for the receiver swaption.

The risk-neutral measure can be used to represent the time- t price of a European payer swaption with notional N , strike K , and tenor structure $\mathcal{T}$ through

$$\text{PS}[t, \mathcal{T}, K, N] = \mathbb{E}^{\mathbb{Q}} \left[D(t, T_\alpha) \left(N (S_{\alpha, \beta}(T_\alpha) - K)^+ \Delta \sum_{i=\alpha+1}^{\beta} P(T_\alpha, T_i) \right) \middle| \mathcal{F}_t \right] = \mathbb{E}^{\mathbb{Q}} [D(t, T_\alpha) \mathcal{P}_{T_\alpha} | \mathcal{F}_t]$$

where $\mathcal{P}_{T_\alpha}$ is the swaption payoff at maturity

$$\mathcal{P}_{T_\alpha} = N \left(1 - P(T_\alpha, T_\beta) - K \Delta \sum_{i=\alpha+1}^{\beta} P(T_\alpha, T_i) \right)^+.$$

As demonstrated in [55], the price can also be expressed by using the T_α -forward measure $\mathbb{Q}^{T_\alpha}$, which uses $P(t, T_\alpha)$ as the numéraire to simplify the pricing formula:

$$\text{PS}[t, \mathcal{T}, K, N] = P(t, T_\alpha) \mathbb{E}^{\mathbb{Q}^{T_\alpha}} [\mathcal{P}_{T_\alpha} | \mathcal{F}_t], \quad (54)$$

This formulation reduces the pricing problem to computing the conditional expectation of a function of future bond prices. Since zero-coupon bond prices under the DTAFNS model are exponential-affine in the factors, the time- T_α bond prices involved in (54) can be fully characterized by the term structure factors X_{T_α} . For dynamics of factors under T -forward measure, refer to Section

3.8.

3.4 Hedging Framework and Replicating Portfolio Construction

We consider a discrete-time hedging framework where hedging portfolio rebalancing dates are $\{0, T_1, \dots, T_\alpha\}$, with $T_i = i\Delta$. The objective is to construct a hedging portfolio to hedge a short position on a European swaption. Without loss of generality, a payer swaption is considered.

To hedge a short position in this option, an agent constructs a dynamic self-financing portfolio comprising M forward-starting payer swaps (which may include the underlying swap) and a money market account that accrues interest at the prevailing short rate. The value of the hedging portfolio at time $t < T_\alpha$ is given by

$$V_t = \phi_t^\top \text{PFS}_t + \psi_t e^{r_{t-1}\Delta}, \quad (55)$$

where $\text{PFS}_t = [\text{PFS}_t^{(1)}, \dots, \text{PFS}_t^{(M)}]^\top$ is the time- t value of the hedging forward-starting payer swaps, ψ_t is the cash amount held at time $t - 1$, which earns interest compounded at rate r_{t-1} , and the hedging strategy is represented by the predictable process $\{(\phi_t, \psi_t)\}_{t=1}^{T_\alpha}$, where $\phi_t = [\phi_t^{(1)}, \dots, \phi_t^{(M)}]^\top$ is the vector of hedge ratios (i.e., the notional value) of the swaps at time $t - 1$.

Under the self-financing condition, the cash holding is determined endogenously as

$$\psi_{t+1} = V_t - \phi_{t+1}^\top \text{PFS}_t - \text{TC}_{t+1}, \quad (56)$$

where TC_{t+1} represents the transaction cost, which is defined as follows:

$$\text{TC}_t = \delta \sum_{k=1}^M \left| \phi_t^{(k)} - \phi_{t-1}^{(k)} \right| \cdot \left| \text{PFS}_{t-1}^{(k)} \right|, \quad (57)$$

where δ denotes a proportional transaction cost rate.

The agent aims to determine an optimal dynamic hedging strategy $\{\phi_t\}_{t=1}^{T_\alpha}$, with the goal of

minimizing the hedging error risk at the swaption maturity T_α .

We refer to the set $\mathcal{X} = \{X_t, t = 0, 1, \dots, T_\beta\}$ as the path of yield curve factors up to time T_β . At each time t , the hedger observes the vector Y_t , which represents the information taken into account to determine their hedging strategy. The composition of Y_t in our setting is detailed subsequently. Based on this input, the hedge ratios are determined by a policy function $\pi_\theta : Y_t \mapsto \phi_{t+1}$, parameterized by a neural network with weights θ

$$\phi_{t+1} = \pi_\theta(Y_t). \quad (58)$$

This approximation allows to represent the sequence of hedging decisions as a parameterized control policy that is learned from simulated market scenarios.

The terminal hedging error is defined as the difference between the swaption payoff and the final hedging portfolio value

$$h_{T_\alpha}(\pi_\theta) = \mathcal{P}_{T_\alpha} - V_{T_\alpha}(\pi_\theta),$$

where the dependence of the terminal portfolio value V_{T_α} on policy parameters θ is made explicit.

A positive value of h_{T_α} indicates a shortfall in the hedging portfolio value—i.e., the hedging strategy failed to offset the liability. The hedger's objective is to find a policy π_θ that minimizes a risk measure $\mathcal{J}$ applied to the hedging error

$$\arg \min_{\theta} \mathcal{J}(\theta), \quad \mathcal{J}(\theta) \equiv \zeta(h_{T_\alpha}(\pi_\theta)) \quad (59)$$

for a given risk measure ζ . This formulation is closely related to the deep hedging framework introduced in [25], where trading strategies are optimized end-to-end using neural networks and Monte Carlo simulations.

3.4.1 Deep Hedging Using Reinforcement Learning

To assess the performance of a given policy, we use Monte Carlo simulation. We generate n independent paths $\mathcal{X}^{(i)}$, $i = 1, \dots, n$, where each path is a realization of the factor process $\{X_t\}_{t=1, \dots, T_\alpha}$. For each path and for a fixed parameter vector θ , we compute the replicating portfolio values using (55), with hedge ratios at each time step being given by (58).

Accordingly, we define the terminal hedging error h_{T_α} as a function of both the market path $\mathcal{X}^{(i)}$ and the hedging strategy generated by the policy π_θ , which we denote by

$$h_{T_\alpha}^{(i)} := h_{T_\alpha}(\pi_\theta; \mathcal{X}^{(i)}).$$

Estimates of the objective function (59) are obtained from the empirical distribution of hedging errors. Denote by n_b the batch size used in training. The mean squared error (MSE) objective is estimated as

$$\hat{\mathcal{J}}_{\text{MSE}}(\theta) = \frac{1}{n_b} \sum_{i=1}^{n_b} \left[h_{T_\alpha}^{(i)} \right]^2, \quad (60)$$

For the downside risk (DR) objective penalizes only under-hedging scenarios and its estimate is

$$\hat{\mathcal{J}}_{\text{DR}}(\theta) = \frac{1}{n_b} \sum_{i=1}^{n_b} \left[\max(h_{T_\alpha}^{(i)}, 0) \right]^2. \quad (61)$$

When the objective is to control tail risk, we consider the CVaR as the performance criterion. Following [101], we define the empirical α -value-at-risk of the hedging error distribution as

$$\hat{v}_\alpha^{n_b} := \inf \left\{ t \in \mathbb{R} \left| \frac{1}{n_b} \sum_{k=1}^{n_b} \mathbf{1}_{\{h_{T_\alpha}^{(k)} \leq t\}} \geq \alpha \right. \right\}. \quad (62)$$

The corresponding CVaR $_\alpha$ is estimated via:

$$\hat{\mathcal{J}}_{\text{CVaR}}(\theta) = \hat{v}_\alpha^{n_b} + \frac{1}{n_b(1-\alpha)} \sum_{i=1}^{n_b} \left[h_{T_\alpha}^{(i)} - \hat{v}_\alpha^{n_b} \right]^+, \quad (63)$$

This loss encourages the policy to minimize the expected shortfall of large hedging errors, improving

robustness highly adverse scenarios.

The optimization problem (59) is tackled using mini-batch stochastic gradient descent (SGD). The parameter vector θ is iteratively updated using:

$$\theta^{(j+1)} = \theta^{(j)} - \eta_j \nabla_{\theta} \hat{\mathcal{J}}(\theta^{(j)}), \quad (64)$$

where the gradient $\nabla_{\theta} \hat{\mathcal{J}}(\theta^{(j)})$ is computed using automatic differentiation in `PyTorch`. Here, the superscript j denotes the iteration index of the SGD procedure. At each iteration, a new batch of Monte Carlo paths $\mathcal{X}$ is generated under the current policy $\theta^{(j)}$, and the risk-based objective $\hat{\mathcal{J}}(\theta^{(j)})$ is estimated from this batch. This procedure ensures that parameter updates are always based on fresh simulations reflecting the evolving policy.

3.4.2 The Rho-Hedging Benchmark

In this section, we introduce the rho-hedging strategy that serves as a benchmark for the learning-based methods developed in this paper.

To hedge the swaption position against market risk, the rho-hedging procedure neutralizes its sensitivity to the key term structure factors $X_t^{(1)}$, $X_t^{(2)}$, or $X_t^{(3)}$ that drive the evolution of interest rates under the DTAFNS model. This is achieved by equating the factor-wise sensitivities (i.e., the rhos) of the swaption and these of the replicating portfolio.

When using a single forward-start swap ($M = 1$) in the replicating portfolio, the sensitivity to a single factor $k \in \{1, 2, 3\}$ is neutralized; the hedge ratio is set to

$$\phi_{t+1}^{(1)} = \frac{\partial \text{PS}_t}{\partial X_t^{(k)}} \bigg/ \frac{\partial \text{PFS}_t^{(1)}}{\partial X_t^{(k)}}.$$

where PS_t denotes the time- t value of the payer swaption.

The choice of the risk factor that is neutralized depends on which risk dimension (e.g., level, slope, or curvature) is deemed most significant for the swaption exposure. As the factor $X_{t-1}^{(1)}$ evolves over time, $\phi_t^{(1)}$ is dynamically updated, resulting in a rho-hedging strategy that rebalances

the portfolio to attempt tracking the swaption's value.

In the case where two swaps are used ($M = 2$), the hedge ratios that neutralize the sensitivity of two of the factors $k \neq j \in \{1, 2, 3\}$ to be chosen are determined by solving the linear system

$$\begin{bmatrix} \frac{\partial \text{PFS}_t^{(1)}}{\partial X_t^{(j)}} & \frac{\partial \text{PFS}_{t-1}^{(2)}}{\partial X_t^{(j)}} \\ \frac{\partial \text{PFS}_t^{(1)}}{\partial X_t^{(k)}} & \frac{\partial \text{PFS}_t^{(2)}}{\partial X_t^{(k)}} \end{bmatrix} \begin{bmatrix} \phi_{t+1}^{(1)} \\ \phi_{t+1}^{(2)} \end{bmatrix} = \begin{bmatrix} \frac{\partial \text{PS}_t}{\partial X_t^{(j)}} \\ \frac{\partial \text{PS}_t}{\partial X_t^{(k)}} \end{bmatrix}.$$

This ensures that the portfolio has a null local sensitivity with respect to both factors j and k . This system of linear equations can be generalized for neutralizing three factors using three swaps. Once the unconstrained hedge ratios have been determined, we impose two practical constraints on the outputs which are detailed in Section 3.5.2.

The price of a PFS is given in Equation (53), and its gradient with respect to factor $k = 1, 2, 3$ is

$$\frac{\partial \text{PFS}_t}{\partial X_t^{(k)}} = -N\Delta \left(B_{T_\alpha-t}^{(k)} P(t, T_\alpha) - B_{T_\beta-t}^{(k)} P(t, T_\beta) - K\Delta \sum_{i=\alpha+1}^{\beta} B_{T_i-t}^{(k)} P(t, T_i) \right). \quad (65)$$

3.5 Market and Model Setup

We use the calibrated DTAFNS model proposed by [42] on end-of-month Canadian spot rate curves from January 1986 to January 2022 (434 months) to simulate the yield curve term structure factors. The calibrated parameters are given in Section 3.9.

Based on this model, we investigate the dynamic hedging of a European payer swaption written on a plain-vanilla forward-starting interest rate swap under realistic market configurations. Specifically, we consider an ATM swaption with a maturity of $T_\alpha = 60$ months (5 years with $\Delta = 1/12$), giving the holder the right to enter into a 5y $\times$ 10y payer swap, i.e. a swap that starts in five years and runs for ten years (a five-years-forward, ten-year-tenor swap), with a fixed rate of $K = 2.5083\%$ and notional of $N = \$1$. The hedging portfolio is rebalanced monthly, using forward-starting payer swaps as hedging instruments. Hedging swaps include the underlying swap and possibly other swaps with different tenors or maturities. The swaption price at time $t = 0$ serves

as the initial value V_0 of the replicating portfolio.

3.5.1 Hedging Network Architecture

For the hedging agent, we employ a four-layer fully connected neural network (FCNN) to approximate the mapping from observed market features to hedge ratios. The width of the respective FCNN hidden layers is $[8, 32, 32, 8]$, followed by a final output layer of size M , where M is the number of forward-starting swaps used in the replicating portfolio. For some input Y_t , the network output is a vector $\phi_{t+1} = \pi_\theta(Y_t)$ representing the hedge ratios for the M swaps at time t before any constraint is imposed. Once the unconstrained hedge ratios have been determined, we impose two constraints on the outputs to reflect risk limits analogous to these applied in practice. The details about the constraints are provided in Section 3.5.2.

The input vector Y_t which includes the current values of the term structure factors X_t , the time-to-maturity $(T_\alpha - t)\Delta$, and the current portfolio value V_t . This input is first normalized using batch normalization. The processed input then passes through a sequence of fully connected layers, each followed by a Mish activation function [90]:

$$x \mapsto x \cdot \tanh(\ln(1 + e^x)),$$

which provides smooth and non-monotonic transformations that are well-suited for financial applications. No dropout is applied, and the network is kept relatively shallow to avoid overfitting given the structured nature of the inputs. We employ a linear activation function for the final layer to determine the weights of the hedging swaps.

The neural network parameters are initialized using the Kaiming uniform method [59], which defines the initial parameter vector $\theta^{(0)}$. It is designed for networks with nonlinear activations and helps maintain stable gradients during training.

Training is performed using the Adam optimizer from [74] with an initial learning rate of 5×10^{-3} . The model is trained for a maximum of 800 epochs. The learning rate is reduced

using a plateau-based scheduler, and early stopping is applied with a patience threshold of 200 epochs to avoid overfitting. Each epoch consists of multiple mini-batches of size $n_b = 2048$, drawn from a simulated dataset of 100,000 Monte Carlo paths. The training objective is to minimize a user-specified risk measure—such as MSE, DR, or CVaR as defined in Section 3.4.1—applied to the terminal hedging error.

3.5.2 Leverage Bounds

The hedging positions outputted by the neural network are unconstrained. In practice, trading desks face binding limits: treasury funding and credit lines limits, clearing margin and liquidity haircuts constraints, counterparty/concentration limits, and book-level risk controls such as gross notional and DV01/PV01 caps¹. To reflect these operational realities and to avoid pro-cyclical “lever up on gains, delever on losses” behavior,² we impose two complementary constraints on our notional positions:

$$\text{(per-leg exposure)} \quad |\phi_{i,t}| |\text{PFS}_{i,t}| \leq L_{\text{per}} (|V_t| + B), \quad \forall i, \quad (66)$$

$$\text{(gross exposure)} \quad \sum_{i=1}^K |\phi_{i,t}| |\text{PFS}_{i,t}| \leq L_{\text{gt}} (|V_t| + B). \quad (67)$$

The absolute value of the $\phi_{i,t}$, the hedging notional position for leg i , and that of the forward-starting swap price $\text{PFS}_{i,t}$, serve as the exposure factor in dollars, consistent with desk practices for notional caps. The first constraint prevents any single leg from dominating the book (mitigating ill-conditioning and concentration risk), while the second controls overall leverage, in line with standard gross-exposure or DV01 guardrails at the portfolio level. The scaling term $(|V_t| + B)$ provides a stabilized risk budget: V_t denotes the current portfolio value (taken in absolute value in implementation), and $B \geq 0$ is a fixed liquidity/capital buffer that (i) avoids vanishing limits when

¹DV01 (Dollar Value of a Basis Point) measures first-order price sensitivity to a 1 bp parallel shift in the yield curve. PV01 (Present Value of 1 bp) is the present value change from a 1 bp change in the fixed rate of a swap

²A constraint scaled by the current hedging portfolio value $|V_t|$ expands risk capacity after gains and contracts it after losses, inducing buy-high/sell-low rebalancing and elevated turnover. We mitigate this by using a stabilized basis $(|V_t| + B)$ in our limits.

$|V_t|$ is small, and (ii) curbs pro-cyclical upsizing when $|V_t|$ spikes. The leverage multipliers L_{per} and L_{gt} are policy parameters reflecting the desk's risk appetite and external constraints (funding, margin, and capital).³ These bounds keep the hedge implementable across liquidity regimes, align with real-world risk controls, and improve numerical stability during learning and backtesting.

Let the (unconstrained) per-leg dollar exposures be

$$e_{i,t} = |\varphi_{i,t}| |\text{PFS}_{i,t}|, \quad i = 1, \dots, M,$$

where $\varphi_{i,t}$ is the unconstrained position on swap i at time t obtained from the neural network. Limits scale with a stabilized basis

$$B_t = |V_t| + B,$$

We enforce a per-leg exposure cap and a gross (portfolio-level) exposure cap:

$$0 \leq x_{i,t} \leq L_{\text{per}} B_t, \quad \sum_{i=1}^M x_{i,t} \leq L_{\text{gt}} B_t,$$

where $x_{i,t} = |\phi_{i,t}| |\text{PFS}_{i,t}|$. At each t , we project $e_t = [e_{1,t}, \dots, e_{M,t}]$ onto the feasible set by

$$x_t^* = \arg \min_{x \in \mathbb{R}^M} \|x - e_t\|_2^2 \quad \text{s.t.} \quad 0 \leq x_i \leq L_{\text{per}} B_t \ (i = 1, \dots, M), \quad \sum_{i=1}^M x_i \leq L_{\text{gt}} B_t.$$

Feasible hedge notionals are then recovered, preserving direction,

$$\phi_{i,t} = \text{sign}(\varphi_{i,t}) \frac{x_{i,t}^*}{|\text{PFS}_{i,t}|}.$$

These constraints align with standard gross-exposure/DV01 guardrails at the portfolio level while ensuring implementable, numerically stable hedges.

We employ this projection on both benchmark and RL weights for certain reasons. Firstly, under the unconstrained problem, we allow the RL agent to freely interact with the environment

³By setting $V_t = V_0$, one can change the dynamic leverage to a static leverage constraint.

and adopt policies in response to market fluctuations. The constraints reduce the smoothness of the objective function and potentially unstable during training. Secondly, we apply the projections to the benchmark to ensure that the constraints are consistent, thereby enhancing the realism and fairness of our evaluation.

3.5.3 Swaption prices and their derivatives

Swaption prices provided by (54) are not available in closed-form. To avoid recomputing estimates of such prices through Monte Carlo simulation every time a price is needed, we pre-compute prices of swaptions offline with a Kolmogorov-Arnold neural network. The methodology is explained in Section 3.10. This approach is also used to compute the first-order sensitivity of swaption prices with respect to term structure risk factors, as also detailed in that section.

3.6 Result Analysis

In this section, we evaluate the performance of the deep hedging strategy and compare it against that of conventional rho-hedging approaches. The assessment is conducted under a comprehensive set of risk and performance metrics. The evaluation is conducted on $N^{(OOS)} = 100,000$ out-of-sample paths generated with the same parameters that are used to generate training set paths.

3.6.1 Hedging Performance Metrics

We assess the distribution of terminal hedging errors using a comprehensive set of risk and performance metrics.

Mean. The *mean* of the hedging errors $h_{T_\alpha}^{(i)}, i = 1, \dots, N^{(OOS)}$ captures the average hedging bias, indicating whether a strategy systematically *overhedges* (negative mean) or *underhedges* (positive mean).

Mean Squared Error (MSE). The MSE, calculated with formula (60) applied to out-of-sample paths, penalizes both over- and underhedging symmetrically and is computed as the average squared deviation from a perfectly replicated payoff.

Downside Risk (DR). Computed through (61) on out-of-sample paths, the downside risk metric quantifies the impact of positive hedging errors, assessing the severity of underhedging losses.

Tail Risk. Tail risk is measured with $\text{CVaR}_{99\%}$, which gives the average value of the worst 1% hedging losses $h_{T_\alpha}^{(i)}$.

Probability of underhedging. We also compute the *probability of underhedging*

$$P(HE > 0) = \frac{1}{N(OOS)} \sum_{i=1}^{N(OOS)} \mathbb{1}_{\{h_{T_\alpha}^{(i)} > 0\}},$$

which reflects the frequency at which the hedging strategy fails to fully replicate the swaption payoff.

Hedging Risk Reduction (HRR). To evaluate the effectiveness of the hedging strategy relative to an unhedged position, we compute the *hedging risk reduction*, defined as

$$\text{HRR} = 1 - \frac{\text{Std}(h_{T_\alpha})}{\text{Std}(h_{T_\alpha}^{\text{unhedged}})},$$

where $h_{T_\alpha}^{\text{unhedged}}$ denotes the hedging error in absence of hedging, i.e. when $\phi \equiv 0$. When no hedging is applied, the swaption premium received at time 0 is deposited in a bank account and accrues risk-free interest until the swaption maturity. The HRR metric represents the percentage reduction in hedging error volatility achieved by hedging.

Turnover. Turnover quantifies the level of trading activity relative to the portfolio's average value over the hedging horizon. For each simulation path i , it is defined as

$$\text{Turnover}_i = \frac{\min(\text{Purchases}_i, \text{Sales}_i)}{\bar{V}_i},$$

where

$$\text{Purchases}_i = \sum_{t=1}^{T_\alpha} \sum_{j=1}^M \mathbf{1}_{\{\Delta\phi_{i,t}^{(j)} > 0\}} \cdot |\Delta\phi_{i,t}^{(j)}| \cdot |\text{PFS}_{i,t}^{(j)}|, \quad \text{Sales}_i = \sum_{t=1}^{T_\alpha} \sum_{j=1}^M \mathbf{1}_{\{\Delta\phi_{i,t}^{(j)} < 0\}} \cdot |\Delta\phi_{i,t}^{(j)}| \cdot |\text{PFS}_{i,t}^{(j)}|,$$

and

$$\bar{V}_i = \frac{1}{T_\alpha} \sum_{t=1}^{T_\alpha} |V_{i,t}|.$$

Here

- $\Delta\phi_{i,t}^{(j)} = \phi_{i,t}^{(j)} - \phi_{i,t-1}^{(j)}$ is the change in the hedge position j at time t for path i .
- $V_{i,t}$ is the portfolio value at time t .

The average turnover across all simulation paths is

$$\text{Turnover} = \frac{1}{N^{(OOS)}} \sum_{i=1}^{N^{(OOS)}} \text{Turnover}_i.$$

Dynamic Tracking Error (DTE). Dynamic Tracking Error measures the pathwise deviation between the hedged portfolio value and the swaption price across the hedging horizon. For each path i , it is defined as

$$\text{DTE}_i = \sqrt{\frac{1}{T_\alpha} \sum_{t=1}^{T_\alpha} (V_{i,t} - S_{i,t})^2},$$

where $S_{i,t}$ is the swaption price at time t for path i . The average DTE across all paths is

$$\text{DTE} = \frac{1}{N^{(OOS)}} \sum_{i=1}^{N^{(OOS)}} \text{DTE}_i.$$

3.6.2 Hedging Performance

This section presents an assessment of the performance of our proposed RL method through Monte Carlo simulation. No transaction costs are considered in this section. The RL method is compared to conventional dynamic rho-hedging approaches. The impact of the number of hedging instruments on hedging performance is assessed, with either one, two or three swaps being included in the hedging portfolio. The hedging single swap case employs the underlying forward-starting swap, which has a start date on the swaption's maturity and whose tenor is that same than that of the swaption. For the two-swap scenario, we augment the hedging portfolio by including an additional 10y×2y swap, i.e. two-year-tenor swap starting in ten years. This second swap is chosen to load primarily on the long end of the curve—capturing movements in long-term rates and in the level and slope components that the underlying hedging instrument alone may not span. Extending further, the three-swap case introduces in addition a 2y×2y swap, which concentrates exposure at the front end of the curve, targeting short-term rate dynamics and improving control of near-dated shocks. Using short tenors and maturities that are far from that of the underlying for the additional hedging swaps allow reducing correlation among hedging legs and protecting against a broader span of risks, yielding a more effective dynamic hedge.

We use the following notation for rho-hedging. $\rho - X^{(i)}$ utilizes a single-swap hedge and neutralizes exposure to the rate of factor i at each rebalancing date. $\rho - (X^{(i)}, X^{(k)})$ hedges rely instead on two swaps and neutralize the exposure to the two factors (i, k) . Lastly, ρ hedging neutralizes exposure to all three factors using three hedging swaps.

Table 3.1 reports performance metrics for RL-based and rho-hedging strategies using one, two, and three swaps on an out-of-sample dataset.

In the single-swap setting, RL methods consistently outperform traditional rho-hedging across the key accuracy and risk measures. The *RL-MSE* strategy achieves the lowest mean squared error (0.000074), highlighting its precision in tracking the target payoff. The *RL-DR* specification provides the strongest downside protection, with the smallest downside risk (0.000010), and the lowest probability of underhedging (0.3092). The *RL-CVaR* approach offers the greatest reduction

Strategy	Mean	MSE	Downside	CVaR _{99%}	P(HE > 0)	HRR	Turnover	DTE
Hedging with One Swap								
RL MSE	-0.0033	0.000074	0.000017	0.0228	0.3220	0.8798	1.1311	0.0049
RL DR	-0.0052	0.000120	0.000010	0.0179	0.3092	0.8551	0.9814	0.0057
RL CVaR	-0.0055	0.000165	0.000016	0.0157	0.3910	0.8251	0.9866	0.0066
$\rho - X^{(1)}$	-0.0046	0.000113	0.000014	0.0185	0.3381	0.8556	1.0587	0.0053
$\rho - X^{(2)}$	-0.0030	0.000204	0.000077	0.0368	0.3971	0.7898	0.8337	0.0106
$\rho - X^{(3)}$	-0.0038	0.000079	0.000016	0.0220	0.2970	0.8790	1.0296	0.0049
Hedging with Two Swaps								
RL MSE	-0.0025	0.000063	0.000017	0.0230	0.3413	0.8860	2.0842	0.0043
RL DR	-0.0087	0.000182	0.000005	0.0147	0.1729	0.8441	1.6550	0.0074
RL CVaR	-0.0081	0.000227	0.000009	0.0131	0.3353	0.8088	1.2354	0.0076
$\rho - (X^{(1)}, X^{(2)})$	-0.0069	0.000148	0.000010	0.0177	0.2386	0.8492	1.5865	0.0065
$\rho - (X^{(1)}, X^{(3)})$	-0.0038	0.001447	0.000602	0.1183	0.4606	0.4296	20.3524	0.0124
$\rho - (X^{(2)}, X^{(3)})$	-0.0096	0.000291	0.000037	0.0313	0.2467	0.7872	1.7024	0.0120
Hedging with Three Swaps								
RL MSE	-0.0014	0.000058	0.000020	0.0236	0.3981	0.8868	3.9573	0.0040
RL DR	-0.0089	0.000187	0.000005	0.0145	0.1629	0.8438	2.5359	0.0076
RL CVaR	-0.0080	0.000205	0.000008	0.0132	0.3028	0.8209	1.2691	0.0075
ρ Hedging	-0.0050	0.000112	0.000012	0.0183	0.3009	0.8600	1.8562	0.0051

Table 3.1: Performance metrics across reinforcement learning and rho-hedging strategies using one, two, and three swaps.

in tail exposure, recording the lowest CVaR_{99%} (0.0157). Furthermore, RL-DR and RL-CVaR strategies maintain moderate turnover, enhancing their practical implementability. By contrast, one-factor rho-hedging delivers weaker control of risk as evidenced by several risk metrics, limiting its effectiveness in this context.

Adding a second swap substantially improves hedging performance for all strategies, but RL retains a clear advantage. The *RL-MSE* model again produces the lowest volatility (0.0076) and mean squared error (0.000063), confirming stable replication accuracy. The *RL-DR* objective minimizes downside risk (0.000005) and yields the lowest probability of underhedging (0.1729), while *RL-CVaR* delivers superior tail performance with a CVaR of 0.0131 at the 99% level.

Adding a third swap to the hedging portfolio only provides marginal additional benefits in terms of performance, both for RL methods and rho-hedging. This indicates that hedging effectiveness

approaches a near-optimal level with two instruments. This can be explained by the high negative correlations between the level factor and the two others as seen in matrix ρ from Section 3.9, which are respectively -0.63 and -0.41 . There is thus insufficient de-correlation between factors to justify the presence of a third hedging instrument. The inability of the third hedging swap to materially improve hedging performance concurs with previous studies suggesting that increasing the number of hedging instruments beyond a certain point yields little incremental benefit and may raise instability ([45, 39, 46]).

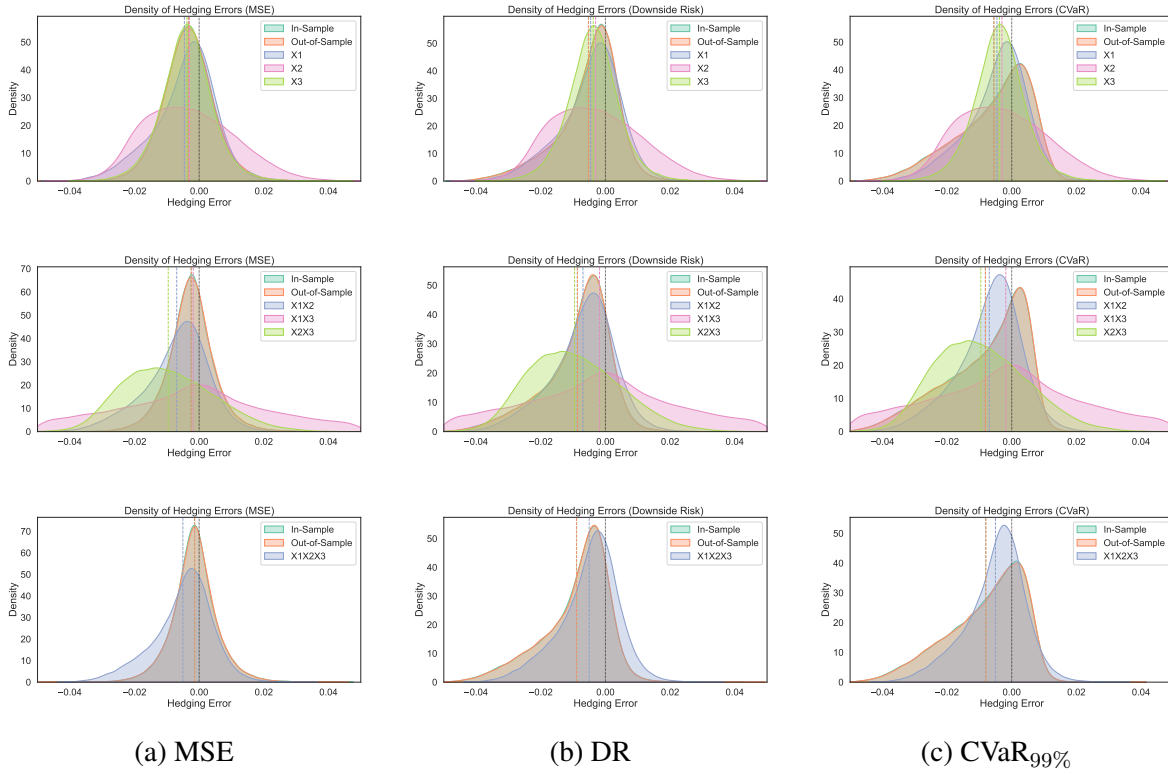


Figure 3.1: Comparison of kernel density estimates of in-sample and out-of-sample hedging errors for RL-based strategies and dynamic rho-hedge strategies (evaluated on the out-of-sample set), across different portfolio compositions. The first row corresponds to portfolios with one swap, the second row to two swaps, and the third row to three swaps.

Figure 3.1 shows kernel density plots comparing the hedging errors of RL and rho-hedging strategies under three risk measures: MSE, DR, and CVaR_{99%}, and using either one, two, or three swaps as hedging instruments. The RL strategies consistently produce tighter, well-centered error distributions, with very little difference between in-sample and out-of-sample results. This pattern

indicates that the RL methods do not overfit the training data. For the MSE objective especially, the RL error curves are symmetric and sharply peak near their mean, indicating accurate and stable hedging. In contrast, the rho-hedging strategies often have wider distributions with heavier tails and shifts in the main peaks—especially when using two or three swaps (such as the $\rho - (X^{(1)}, X^{(3)})$ setup). This implies more frequent large errors and some bias in replication. Under the DR and CVaR objectives, RL methods again show tighter error distributions, confirming that they better manage risks related to extreme losses and downside fluctuations. Meanwhile, the rho-hedging approaches produce broader, less stable error profiles.

Overall, RL-based hedging consistently outperforms rho-hedging by delivering superior risk mitigation performance.

3.6.3 Residual Sensitivity Analysis

This section analyzes whether the outperformance of RL hedging strategies is partially due to systematic exposure to some risk factors in order to harvest associated risk premia. We conduct a residual exposure sensitivity analysis. Define, for any time step t , the residual vector, $\xi_t = [\xi_t^{(1)}, \xi_t^{(2)}, \xi_t^{(3)}]^\top$:

$$\xi_t = A_t \phi_{t+1} - \sigma_t,$$

where σ_t is the vector of the swaptions' sensitivities to the underlying risk factors, and A_t is a $M \times 3$ matrix of sensitivities of the available swaps in the replicating portfolio, where M is the number of swaps in the replicated portfolio:

$$\left(\sigma_t\right)_i = \frac{\partial \text{PS}_t}{\partial X_t^{(i)}}, \quad \text{and} \quad \left(A_t\right)_{j,i} = \frac{\partial \text{PFS}_t^{(j)}}{\partial X_t^{(i)}}, \quad \text{for } i = 1, 2, 3, \quad j = 1, \dots, M.$$

By plotting the time series of the mean value of each component $\xi_t^{(k)}$ across all paths, we can identify whether there is a systematic bias in any direction. If $\xi_t^{(k)}$ fluctuates around zero with low variance, the RL agent behaves similarly to a traditional rho-hedger. Conversely, if $\xi_t^{(k)}$ consistently remains above (below) zero, it indicates that the RL agent systematically positions itself to benefit

from an increase (decrease) in the value of factor $X^{(k)}$. This analysis helps to assess whether the observed outperformance arises at least partially from directional exposure. To extend the analysis, we examine the agent's behavior along two sets of paths: those that ultimately result in either (i) a positive payoff or (ii) a zero payoff. Focusing on these scenarios separately provides insight into how the agent adjusts its actions based on the option moneyness as the portfolio approaches maturity.

The three factors, corresponding respectively to level ($X^{(1)}$), slope ($X^{(2)}$), and curvature ($X^{(3)}$) of the yield curve, are associated with distinct risk premia. The level factor is associated with long-term interest rates and inflation expectations. This factor is linked to the term premium: investors typically require compensation for holding long-duration bonds. The slope (factor $X^{(2)}$) is associated with steepening/flattening of the yield curve and reflects the difference between short- and long-term rates. It is linked to carry/roll-down trade premium. The curvature (factor $X^{(3)}$) represents medium-term deviations, often tied to intermediate-term economic uncertainty or changing term premia. It is associated with convexity and volatility-of-volatility risk, where investors often demand compensation for bearing curvature exposure.

In our model, we can think about risk premia by comparing the long-run averages under the physical and risk-neutral measures ($\theta^{\mathbb{P}} - \theta^{\mathbb{Q}}$). For the slope and curvature factors, this difference points to premia of about -3.3% and -2.6% , respectively. For the level factor, however, things are different: since it follows a unit root under $\mathbb{Q}$, its long-term mean isn't well defined. Instead, its expected value just stays fixed at $X_0^{(1)} = -0.0312$. Because of this, we can't pin down a structural risk premium for the level factor in the same way, though its expected contribution under $\mathbb{P}$ does turn out to be positive. Interpreting these results, we can see that taking systematic long positions in slope or curvature is effectively paying a negative premium—like buying insurance against bad states. Cutting back on those exposures, in turn, reflects efforts to save on costs. On the other hand, holding level exposure is more like collecting its positive expected payoff, where trimming it usually reflects a desire to manage risk rather than making a simple directional bet.

Because the level factor follows an integrated random walk under $\mathbb{Q}$, it is non-stationary. This

makes it tricky to define a long-term premium for the level factor, since its conditional expected long-run value under $\mathbb{Q}$ is always just the current level $X_t^{(1)}$. Even though this means a clear risk premium for the level factor cannot be established, RL agents still take positions in it. They do so for two main reasons: first, the level factor directly determines the moneyness of the swaption and therefore its likelihood of paying off; and second, it drives the majority of the variance in hedging error.

In Figure 3.2, we track the residual exposure to the three term structure factors of a swaption position hedged with two swaps. Results are presented for the RL agents using the MSE, DR, and CVaR_{99%} risk measures as objectives, and for rho-hedging.

The MSE agent responds symmetrically to hedging errors, aiming to minimize average deviations rather than guard against tail risks. When the payoff is likely to turn positive, it builds a long position in the level factor, benefiting from its positive expected contribution under $\mathbb{P}$. At the same time, it may take a modest long position in the slope factor early on, consistent with capturing short-term roll-down effects from a steep initial curve, even though the long-run premium on the slope factor is negative. Exposure to curvature remains negligible throughout. In contrast, when the trade is unlikely to finish in-the-money as maturity approaches, the agent reduces its level exposure and may even go short, giving up potential upside in order to balance squared errors across different states. Its slope exposure is largely unchanged across payoff scenarios, while curvature consistently remains minimal. Overall, the MSE strategy is simple and focused: it actively manages level risk while treating slope and curvature as secondary, avoids systematic insurance-style trades, and instead concentrates on minimizing mean-squared hedging error. Consistent with the all-paths residual analysis, the agent's exposures are centered around zero, reflecting its goal of balancing errors rather than reshaping tails.

The DR agent adjusts its exposures dynamically based on the probability of the swaption finishing in-the-money. When the swaption is in-the-money, it moves into short positions in the level, slope, and curvature factors, effectively over-hedging to eliminate residual shortfalls and lock in accumulated gains. When the swaption is out-of-the-money, the strategy reverses: the

agent builds long positions in the level factor to improve the chances of recovery and adds long exposure to slope and curvature as maturity nears, using these factors to limit the risk of ending with a large shortfall. In this way, the DR strategy systematically accepts negative premia in order to reduce under-hedging across many states, behaving like an investor who willingly sacrifices carry to preserve downside protection.

The CVaR agent is the most tail-focused of the three, concentrating its adjustments on scenarios that generate large under-hedges. When the swaption is in-the-money, it adopts a defensive posture: cutting level exposure and taking short positions in slope and curvature to over-hedge against extreme losses and safeguard realized gains. When the swaption is out-of-the-money, it shifts to a more opportunistic stance, going long on the level factor to improve the chance of recovery and adding targeted slope and curvature exposures near maturity—even at the cost of paying negative premia. As a result, the CVaR strategy is highly state-dependent: defensive when ahead and tactical when behind, with positions designed explicitly to reshape the distribution of terminal hedging errors in the worst-case scenarios.

The three reinforcement learning agents display distinct yet complementary hedging styles. The MSE agent is premium-neutral, focusing on actively managing level risk while leaving slope and curvature largely untouched, with the objective of minimizing average squared errors rather than reshaping the tails of the distribution. The DR agent is more insurance-oriented: it systematically over-hedges when the swaption is in-the-money, paying negative premia on slope and curvature to suppress shortfalls, while shifting long on level when out-of-the-money to preserve recovery potential. The CVaR agent is the most tail-sensitive, cutting level exposure and turning short on slope and curvature to eliminate extreme losses when ahead, but deliberately re-risking when behind by building level exposure and selectively adding slope and curvature positions. Taken together, these strategies illustrate how RL agents move beyond static hedging rules: they dynamically learn when to harvest positive contributions, when to pay for protection, and how to tailor exposures state by state in order to balance upside with disciplined downside control.

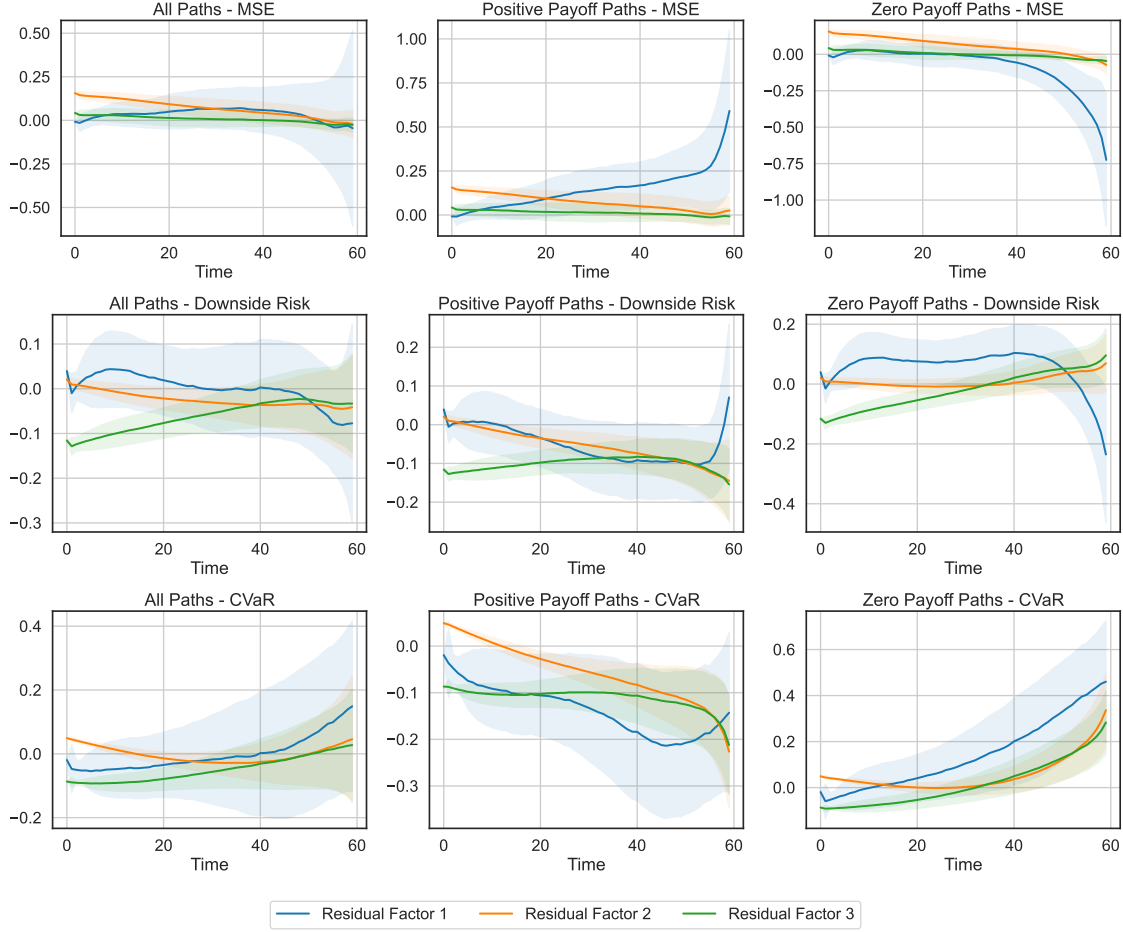


Figure 3.2: Residuals for 2-Swap Portfolio – Multiple Objectives. Shaded areas represent scaled standard deviation bands ($0.5 \times \text{std}$) for visual clarity

3.6.4 Feature Importance

To understand what drives RL hedging strategies, it is important to evaluate how much influence each model input has on the rebalancing decisions being applied. Analyzing feature importance allows us to see which of the state variables play a bigger role in hedging position adjustments, improving both transparency and interpretability. We assess the contribution of the term structure factors, namely the level, slope, and curvature of the yield curve, the portfolio’s value, and the time-to-maturity, and examine how they shape the model’s hedging policy.

In our setting, we apply SHapley Additive exPlanations (SHAP) to the hedging portfolio positions generated by the RL policy. This framework follows the approach of Štrumbelj and Kononenko [102], later formalized by Lundberg and Lee [86]. SHAP values decompose each

hedging decision into contributions from the various state variables (i.e. features) by measuring each feature's marginal contribution. Importantly, these values are obtained without retraining the policy. Instead, the trained model is treated as a black box, and predictions are compared across subsets of features replaced by background samples. We refer readers to the above-mentioned references for more details about this approach and its methodology.

The resulting beeswarm plots shown in Figure 3.3 provide an intuitive way to interpret these attributions. The horizontal axis represents the SHAP value, i.e., the impact of a feature on the hedging position: positive values push the position upward, while negative values push it downward. Each dot corresponds to one feature observation. The color encodes the magnitude of the feature value, with red indicating high values and blue indicating low values, revealing whether the hedge reacts positively or negatively to value of that feature. The vertical dispersion of dots reflects the overall importance of a feature across all scenarios: features with a wider spread exert stronger influence on the RL model's hedging decisions. Taken together, these elements reveal not only which features matter most, but also how their high or low realizations drive hedging position adjustments.

The case of hedging portfolios composed of two hedging swaps is considered. Under the MSE objective, the portfolio value is the dominant driver of the first swap ($5y \times 10y$), with a direct effect: as the swaption's mark-to-market rises (with which the portfolio value is highly correlated), the hedge position increases. This reflects that the first swap is the natural replication instrument for the swaption payoff, so its position mirrors the swaption's valuation and its sensitivity to the level of rates. The negative influence of the level and curvature factors does not mean a simple reduction in exposure, but rather that increases in rates or convexity lead to more negative hedge positions to counteract the swaption's reduced sensitivity. For the second swap ($10y \times 2y$), the pattern differs: the portfolio value exerts an opposite effect, while the level factor has a strong direct influence. This indicates that the auxiliary swap is used primarily to rebalance residual exposures that are not fully hedged by the main instrument. Time-to-maturity has a smaller, negative effect, reflecting modest adjustments as expiry approaches.

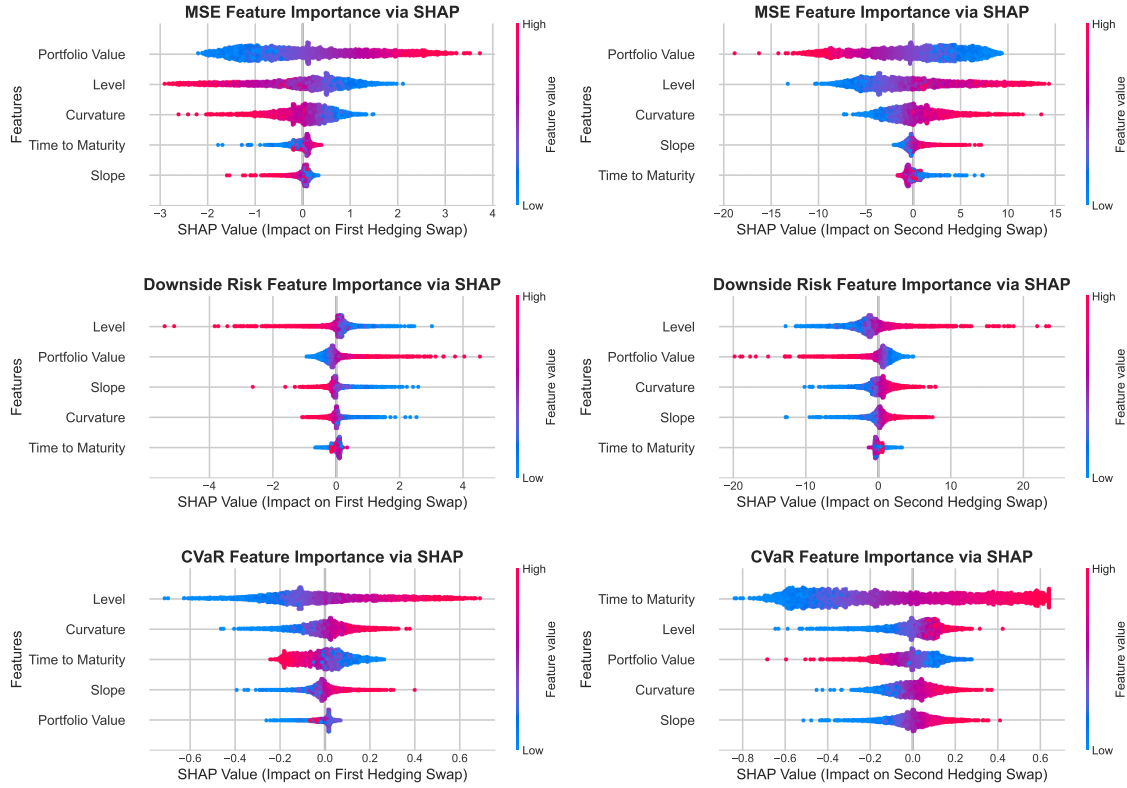


Figure 3.3: Shapley feature importance of hedge swap weights for two hedging strategies (MSE, DR, CVaR).

When minimizing downside risk, the first swap ($5y \times 10y$) hedge is driven primarily by the level factor, with a negative effect, meaning the strategy increases short positions when rates rise to contain under-hedge in adverse scenarios. The portfolio value plays a secondary, direct role. This reflects a more defensive stance: the main hedge turns more negative as rates rise sharply, even if the portfolio value increases. For the second swap ($10y \times 2y$), the level is again the most important feature, while curvature and slope also contribute materially, suggesting that the second swap is used to fine-tune exposures against more complex curve risks. Compared to MSE, the DR hedge distributes sensitivity more evenly across all yield curve factors, reflecting its emphasis on reducing under-hedges broadly rather than simply minimizing variance.

Under CVaR, positions on the first swap ($5y \times 10y$) remain driven by the level and curvature factors, but the effects are smoother and smaller. This indicates that the primary hedge continues to follow the swaption's structural sensitivities, but without the aggressive adjustments observed

under MSE or DR. In particular, the CVaR agent uses Swap 1 to manage curvature and maturity risk in tail scenarios, consistent with its objective of targeting extreme outcomes. The second swap (10y×2y) is dominated by time-to-maturity, with a direct effect, meaning its hedge position increases as the swaption approaches expiry. This swap is therefore used more lightly and primarily for maturity-driven adjustments.

In summary, the first swap (5y×10y) is always the main replication instrument, with its weight primarily tied to the portfolio value and the level factor. This reflecting its role in replicating the swaption payoff. The second swap (10y×2y) plays a complementary role, absorbing residual risks. Under the MSE objective, it helps control variance across different curve states. Under the DR objective it balances slope and curvature exposures to avoid losses in adverse outcomes. Under the CVaR objective, it acts as a tail-risk hedge, especially near expiry. This division of roles highlights how reinforcement learning tailors hedge usage across instruments depending on the risk objective.

3.6.5 Impact of Transaction Costs

Transaction costs represent an essential component of practical hedging strategies, as frequent portfolio rebalancing can erode the profitability of otherwise risk-efficient hedges. The goal of this section is to evaluate how trading frictions affect the performance of RL-based swaption hedging policies relative to traditional rho-hedging benchmarks. Incorporating transaction costs is particularly important because strategies optimized under frictionless assumptions may generate excessive turnover, leading to inflated implementation costs in real markets. Recall that we model transaction costs as a proportional penalty on changes in notional exposure at each rebalancing step, as indicated by (57), and deduct these costs from the portfolio value. This setup allows us to quantify the trade-off between risk reduction and cost efficiency, and to assess whether reinforcement learning methods trained in the presence of transaction costs adaptively limit trading activity in order to preserve performance. In our analysis, hedging positions from the rho-hedging benchmarks are indirectly impacted through the imposed dynamics trading bounds that are detailed in Section 3.5.2. In our analysis, we set the transaction cost multiplier in equation (57) to $\delta = 0.001$ (10 basis points),

applied proportionally to the absolute change in notional exposure at each rebalancing step. to quantify the net effect of trading frictions.

Strategy	Mean	MSE	Downside	CVaR 99%	P(HE > 0)	HRR	Turnover	DTE	TC
Hedging with One Swap									
RL MSE	-0.0030	0.000090	0.000028	0.0278	0.3449	0.8638	1.0556	0.0061	0.000145
RL DR	-0.0050	0.000116	0.000010	0.0179	0.3197	0.8560	0.9787	0.0056	0.000129
RL CVaR	-0.0054	0.000167	0.000016	0.0158	0.3993	0.8234	0.9880	0.0066	0.000127
$\rho - X^{(1)}$	-0.0044	0.000126	0.000024	0.0243	0.3370	0.8442	0.9589	0.0067	0.000133
$\rho - X^{(2)}$	-0.0022	0.000230	0.000096	0.0382	0.4399	0.7738	0.7793	0.0114	0.000128
$\rho - X^{(3)}$	-0.0035	0.000095	0.000026	0.0271	0.3159	0.8631	0.9381	0.0062	0.000132
Hedging with Two Swaps									
RL MSE	-0.0022	0.000061	0.000018	0.0229	0.3544	0.8868	2.1874	0.0043	0.000246
RL DR	-0.0085	0.000180	0.000005	0.0152	0.1728	0.8442	1.5359	0.0074	0.000188
RL CVaR	-0.0079	0.000215	0.000009	0.0133	0.3334	0.8132	1.2147	0.0074	0.000156
$\rho - (X^{(1)}, X^{(2)})$	-0.0067	0.000146	0.000010	0.0179	0.2477	0.8489	1.5893	0.0064	0.000202
$\rho - (X^{(1)}, X^{(3)})$	-0.0016	0.001444	0.000667	0.1219	0.4891	0.4280	20.4239	0.0124	0.002075
$\rho - (X^{(2)}, X^{(3)})$	-0.0093	0.000286	0.000038	0.0316	0.2503	0.7875	1.7034	0.0119	0.000212
Hedging with Three Swaps									
RL MSE	-0.0012	0.000056	0.000021	0.0238	0.4073	0.8882	4.3337	0.0040	0.000398
RL DR	-0.0087	0.000185	0.000005	0.0153	0.1743	0.8425	1.6755	0.0075	0.000197
RL CVaR	-0.0079	0.000216	0.000009	0.0130	0.3277	0.8132	1.2508	0.0075	0.000155
ρ Hedging	-0.0048	0.000110	0.000013	0.0186	0.3113	0.8597	1.8604	0.0051	0.000205

Table 3.2: Performance metrics with transaction costs ($\delta = 10$ bps) across reinforcement learning and ρ -hedging strategies using one, two, and three swaps.

Table 3.2 reports the performance metrics with transaction costs for the various RL and rho-hedging strategies. As expected, RL strategies optimized under the MSE criterion result in hedges deliver the highest accuracy in terms of terminal hedging error MSE. However, in the presence of transaction costs, an MSE objective can have an undesirable effect on the behavior of the hedging agent. Indeed, when the portfolio value V is high, the agent can increase turnover in order to generate losses and reduce tracking error. This behavior results in substantial implementation costs. In contrast, RL strategies trained under DR and CVaR objectives adopt more restrained trading patterns in comparison to RL-MSE, meaning they adjust hedge positions less frequently and with smaller notional shifts. These agents naturally reduce turnover, as seen in the Turnover column of

Table 3.2, which keeps transaction costs moderate while still preserving the intended risk-protection benefits. As a result, they offer greater cost- efficiency in practice.

When comparing the different objectives, a clear trade-off emerges. MSE-based hedges provide superior error reduction in terms of MSE, but they require more frequent rebalancing. DR-based hedges achieve the lowest downside exposure while incurring the smallest transaction costs, and CVaR-based hedges offer a balanced solution by delivering robust tail-risk protection at manageable cost levels. Notably, the reinforcement learning framework exhibits adaptability in the presence of trading frictions. Once transaction costs are introduced, RL agents automatically adjust their trading frequency downward, in sharp contrast to traditional rho-hedging benchmarks, which remain rigid and often generate very high costs, particularly in multi-swap settings.

Overall, the findings demonstrate that RL-based approaches not only outperform conventional rho-hedging in frictionless markets but also preserve this advantage when trading frictions are considered. By implicitly learning turnover-minimizing behaviors, even without explicit instructions, RL policies produce hedges that are both risk-efficient and cost-effective. This underscores their practical relevance for real-world swaption hedging.

3.6.6 Market Sensitivity Analysis

In this section, we examine the robustness of the hedging strategies to model misspecification. Specifically, we consider separately perturbations on two key sets of parameters of the yield curve factor process in (49): the mean-reversion speed parameters $\kappa^{\mathbb{P}}$ and the long-run means $\theta^{\mathbb{P}}$. We then evaluate out-of-sample hedging performance of agents trained on original parameters when hedging on paths generated with perturbed parameters. Benchmarks also rely on hedging positions calculated with original parameters. To introduce stress, we define perturbed parameter as

$$\kappa' = c_{\kappa} \kappa^{\mathbb{P}}, \quad \theta' = c_{\theta} \theta^{\mathbb{P}},$$

with scaling constants set to $c_{\kappa} = c_{\theta} = 1.20$.

Table 3.3 and Table 3.4 report the performance of hedging strategies under perturbations respectively to the mean-reversion speed parameters $\kappa^{\mathbb{P}}$, and to the long-run mean parameters $\theta^{\mathbb{P}}$.

The impact of parameter misspecification differs between the two perturbation scenarios. When $\kappa^{\mathbb{P}}$ is perturbed, the deterioration in hedging performance is moderate. The RL strategies maintain relatively low hedging error MSE and controlled tail risk, whereas the rho-hedges show higher sensitivity to parameter perturbation. Errors in the speed of mean-reversion mainly affect the persistence of market shocks, which RL policies are able to absorb without large losses in efficiency.

By contrast, perturbations to $\theta^{\mathbb{P}}$ lead to more pronounced degradation of performance across all metrics. Shifting the long-run mean introduces a structural drift in the factor dynamics, which results in larger MSE, downside risk and CVaR_{99%}. Both RL and ρ -hedges deteriorate under these conditions, but the RL agents still have higher performance than the traditional first-order sensitivity strategies.

While all methods deteriorate under $\theta^{\mathbb{P}}$ -shifts, RL strategies consistently deliver lower tail risk, more stable turnover, and better hedging risk reduction than rho-hedges. Overall, the evidence indicates that RL-based hedging provides superior robustness to parameter misspecification.

3.7 Zero-Coupon price

As shown in [42], the time- t price of a risk-free zero-coupon bond paying one dollar on maturity $T > t$ is, under the such model,

$$P(t, T) = A_{\tau} \exp \left[-\Delta B_{\tau}^{\top} X_t \right],$$

where $\tau = T - t$, $B_{\tau} = \left[B_{\tau}^{(1)}, B_{\tau}^{(2)}, B_{\tau}^{(3)} \right]^{\top}$ and

$$B_{\tau}^{(1)} = \tau, \quad B_{\tau}^{(2)} = \frac{1 - (1 - \lambda)^{\tau}}{\lambda}, \quad B_{\tau}^{(3)} = \frac{1 - (1 - \lambda)^{\tau-1}}{\lambda} - (\tau - 1)(1 - \lambda)^{\tau-1},$$

Strategy	Mean	MSE	Downside	CVaR 99%	P(HE > 0)	HRR	Turnover	DTE	TC
Hedging with One Swap									
RL MSE	-0.0030	0.000090	0.000028	0.0278	0.3449	0.8638	1.0556	0.0061	0.000145
RL DR	-0.0050	0.000126	0.000017	0.0226	0.3192	0.8481	0.9265	0.0071	0.000132
RL CVaR	-0.0054	0.000166	0.000020	0.0192	0.3706	0.8235	0.9370	0.0080	0.000130
$\rho - X^{(1)}$	-0.0044	0.000126	0.000024	0.0243	0.3370	0.8442	0.9589	0.0067	0.000133
$\rho - X^{(2)}$	-0.0022	0.000230	0.000096	0.0382	0.4399	0.7738	0.7793	0.0114	0.000128
$\rho - X^{(3)}$	-0.0035	0.000095	0.000026	0.0271	0.3159	0.8631	0.9381	0.0062	0.000132
Hedging with Two Swaps									
RL MSE	-0.0022	0.000079	0.000029	0.0281	0.3614	0.8712	1.8647	0.0055	0.000222
RL DR	-0.0090	0.000194	0.000007	0.0173	0.1816	0.8406	1.4947	0.0098	0.000197
RL CVaR	-0.0081	0.000218	0.000010	0.0159	0.2984	0.8163	1.1463	0.0096	0.000159
$\rho - (X^{(1)}, X^{(2)})$	-0.0069	0.000163	0.000015	0.0208	0.2614	0.8397	1.4860	0.0087	0.000203
$\rho - (X^{(1)}, X^{(3)})$	-0.0015	0.001466	0.000695	0.1249	0.4819	0.4285	18.7467	0.0130	0.002073
$\rho - (X^{(2)}, X^{(3)})$	-0.0102	0.000308	0.000036	0.0322	0.2320	0.7868	1.6080	0.0138	0.000213
Hedging with Three Swaps									
RL MSE	-0.0011	0.000070	0.000032	0.0284	0.4142	0.8763	3.6996	0.0049	0.000360
RL DR	-0.0092	0.000198	0.000006	0.0170	0.1748	0.8396	1.6135	0.0100	0.000206
RL CVaR	-0.0081	0.000220	0.000010	0.0157	0.3050	0.8149	1.1857	0.0096	0.000159
ρ Hedging	-0.0048	0.000126	0.000021	0.0232	0.3175	0.8487	1.7174	0.0068	0.000205

Table 3.3: Performance metrics under $\kappa^{\mathbb{P}}$ -shocks with transaction costs $\delta = 0.001$ (10 bps) across reinforcement learning and ρ -hedging strategies using one, two, and three swaps.

$$\log A_{\tau} = -\Delta\theta_2^{\mathbb{Q}}(B_{\tau}^{(1)} - B_{\tau}^{(2)}) + \Delta\theta_3^{\mathbb{Q}}B_{\tau}^{(3)} + \frac{1}{2}\Delta^2v_{\tau},$$

with

$$v_{\tau} = \left(\sum_{i=1}^3 \sum_{j=1}^3 v_{\tau}^{(i,j)} \right),$$

$$v_{\tau}^{(1,1)} = \sum_{1,1}^2 \frac{\tau(\tau-1)(2\tau-1)}{6},$$

$$v_{\tau}^{(2,2)} = \frac{\sum_{2,2}^2}{\lambda^2} \left(\tau - 2 \left[\frac{1 - (1-\lambda)^{\tau}}{\lambda} \right] + \frac{1 - (1-\lambda)^{2\tau}}{1 - (1-\lambda)^2} \right),$$

$$v_{\tau}^{(3,3)} = \mathbb{1}_{\{\tau > 1\}} \frac{\sum_{3,3}^2}{\lambda^2} \left[\tau - 2 + \zeta_0((1-\lambda)^2, \tau-1) + \lambda^2 \zeta_2((1-\lambda)^2, \tau-1) \right.$$

$$\left. - 2\zeta_0((1-\lambda), \tau-1) - 2\lambda\zeta_1((1-\lambda), \tau-1) + 2\lambda\zeta_1((1-\lambda)^2, \tau-1) \right],$$

Strategy	Mean	MSE	Downside	CVaR 99%	P(HE > 0)	HRR	Turnover	DTE	TC
Hedging with One Swap									
RL MSE	-0.0026	0.000180	0.000066	0.0353	0.4386	0.8018	0.9358	0.0114	0.000147
RL DR	-0.0042	0.000214	0.000051	0.0307	0.4380	0.7891	0.8271	0.0123	0.000135
RL CVaR	-0.0045	0.000239	0.000053	0.0264	0.4177	0.7773	0.8314	0.0126	0.000132
$\rho - X^{(1)}$	-0.0037	0.000201	0.000055	0.0297	0.4303	0.7937	0.8955	0.0119	0.000139
$\rho - X^{(2)}$	-0.0006	0.000336	0.000160	0.0419	0.5101	0.7240	0.7511	0.0129	0.000136
$\rho - X^{(3)}$	-0.0031	0.000174	0.000054	0.0323	0.4307	0.8070	0.8796	0.0116	0.000139
Hedging with Two Swaps									
RL MSE	-0.0019	0.000178	0.000071	0.0362	0.4655	0.8226	1.7183	0.0109	0.000239
RL DR	-0.0070	0.000263	0.000038	0.0293	0.3811	0.8038	1.3696	0.0143	0.000200
RL CVaR	-0.0065	0.000270	0.000040	0.0249	0.3727	0.7974	1.0310	0.0139	0.000163
$\rho - (X^{(1)}, X^{(2)})$	-0.0053	0.000230	0.000048	0.0294	0.3967	0.8093	1.4400	0.0133	0.000216
$\rho - (X^{(1)}, X^{(3)})$	-0.0009	0.001515	0.000752	0.1287	0.4909	0.4775	17.8646	0.0160	0.002219
$\rho - (X^{(2)}, X^{(3)})$	-0.0090	0.000357	0.000064	0.0415	0.2767	0.7770	1.5641	0.0183	0.000224
Hedging with Three Swaps									
RL MSE	-0.0013	0.000167	0.000071	0.0353	0.4871	0.8273	3.3148	0.0105	0.000383
RL DR	-0.0072	0.000264	0.000037	0.0291	0.3702	0.7808	1.4679	0.0144	0.000207
RL CVaR	-0.0066	0.000277	0.000042	0.0248	0.3788	0.7945	1.0717	0.0140	0.000162
ρ Hedging	-0.0043	0.000206	0.000051	0.0298	0.4192	0.8162	1.6140	0.0123	0.000216

Table 3.4: Performance metrics under $\theta^{\mathbb{P}}$ -shocks with transaction costs $\delta = 0.001$ (10 bps) across reinforcement learning and ρ -hedging strategies using one, two, and three swaps.

$$\begin{aligned}
v_{\tau}^{(1,2)} &= v_{\tau}^{(2,1)} = \rho_{1,2} \Sigma_{1,1} \Sigma_{2,2} \frac{1}{\lambda} \left(\frac{\tau(\tau-1)}{2} - \zeta_1((1-\lambda), \tau) \right), \\
v_{\tau}^{(1,3)} &= v_{\tau}^{(3,1)} = \mathbf{1}_{\{\tau > 1\}} \rho_{1,3} \Sigma_{1,1} \Sigma_{3,3} \frac{1}{\lambda} \left[\frac{\tau(\tau-1)}{2} - 1 - \zeta_0((1-\lambda), \tau-1) \right. \\
&\quad \left. - (\lambda+1) \zeta_1((1-\lambda), \tau-1) - \lambda \zeta_2((1-\lambda), \tau-1) \right], \\
v_{\tau}^{(2,3)} &= v_{\tau}^{(3,2)} = \mathbf{1}_{\{\tau > 1\}} \rho_{2,3} \Sigma_{2,2} \Sigma_{3,3} \left(\frac{\tau-2 - (2-\lambda) \zeta_0((1-\lambda), \tau-1) + (1-\lambda) \zeta_0((1-\lambda)^2, \tau-1)}{\lambda^2} \right. \\
&\quad \left. + \frac{-\zeta_1((1-\lambda), \tau-1) + (1-\lambda) \zeta_1((1-\lambda)^2, \tau-1)}{\lambda} \right), \\
\zeta_0(r, \tau) &\equiv \sum_{u=1}^{\tau-1} r^u = \frac{r - r^{\tau}}{1-r},
\end{aligned}$$

$$\zeta_1(r, \tau) \equiv \sum_{u=1}^{\tau-1} ur^u = \frac{r - \tau r^\tau + (\tau - 1)r^{\tau+1}}{(1 - r)^2},$$

$$\zeta_2(r, \tau) \equiv \sum_{u=1}^{\tau-1} u^2 r^u = \frac{-(\tau - 1)^2 r^{\tau+2} + (2\tau^2 - 2\tau - 1)r^{\tau+1} - \tau^2 r^\tau + r^2 + r}{(1 - r)^3}.$$

3.8 Factor Dynamics under the T -forward Measure

Results from this appendix are drawn from [55]. The T -forward measure, denoted by $\mathbb{Q}^T$, is defined using the price of a zero-coupon bond maturing at time T as the numéraire. The Radon–Nikodym derivative for changing from the risk-neutral measure $\mathbb{Q}$ to $\mathbb{Q}^T$ is given by:

$$\frac{d\mathbb{Q}^T}{d\mathbb{Q}} = \frac{D(0, T)}{P(0, T)}.$$

Let $\tau = T - t$. Under the T -forward measure, the innovation process satisfies:

$$Z_{t+1}^T = Z_{t+1}^{\mathbb{Q}} + \Delta \Sigma B_{\tau-1},$$

and the conditional distribution of Z_{t+1}^T given $\mathcal{F}_t$ is Gaussian with zero mean and covariance matrix ρ . As shown in Proposition 2.1 and Corollary 2.1 of [55], the sequence $\{Z_j^T\}_{j=1}^T$ is conditionally independent under $\mathbb{Q}^T$.

Using this, the dynamics of the term structure factors under the T -forward measure can be written in the affine form:

$$X_{t+1} = X_t - \eta_t^T + \kappa^T(\theta^T - X_t) + \Sigma Z_{t+1}^T,$$

where the parameters are defined as:

$$\theta^T = \theta^{\mathbb{Q}}, \quad \kappa^T = \kappa^{\mathbb{Q}}, \quad \eta_t^T = \Delta \Sigma \rho \Sigma B_{\tau-1}.$$

Finally, Proposition 2.2 in the same paper states that for any $t + n \leq T$, the future factor values

X_{t+n} , conditionally on $\mathcal{F}_t$, follow a multivariate Gaussian distribution under $\mathbb{Q}^T$ with explicitly computable mean vector and covariance matrix whose formulas are given in [55]. This tractability plays a crucial role in pricing and hedging derivatives such as swaptions under the DTAFNS model.

3.9 Model Parameter Setup

We use the calibrated DTAFNS model parameters proposed in [42]. They perform a joint estimation of parameters from measures $\mathbb{P}$ and $\mathbb{Q}$, which gives the following estimates:

- **Shape parameter:** $\lambda = 0.0233$.

- **Long-term means:**

$$\theta^{\mathbb{P}} = \begin{bmatrix} 0.0000 \\ 0.0301 \\ 0.0505 \end{bmatrix}, \quad \theta^{\mathbb{Q}} = \begin{bmatrix} 0.0000 \\ 0.0633 \\ 0.0766 \end{bmatrix}.$$

- **Correlation matrix of innovations:**

$$\rho = \begin{bmatrix} 1.0000 & -0.6303 & -0.4097 \\ -0.6303 & 1.0000 & 0.2993 \\ -0.4097 & 0.2993 & 1.0000 \end{bmatrix}.$$

- **Volatilities (diagonal of Σ):**

$$\text{diag}(\Sigma) = \begin{bmatrix} 0.0027 \\ 0.0045 \\ 0.0070 \end{bmatrix}.$$

- **Speed of mean-reversion (under $\mathbb{P}$):**

$$\kappa^{\mathbb{P}} = \begin{bmatrix} 0.0075 & 0.0000 & 0.0000 \\ 0.0000 & 0.0288 & -0.0233 \\ 0.0000 & 0.0000 & 0.0354 \end{bmatrix}.$$

- **Speed of mean-reversion (under $\mathbb{Q}$):**

$$\kappa^{\mathbb{Q}} = \begin{bmatrix} 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.0233 & -0.0233 \\ 0.0000 & 0.0000 & 0.0233 \end{bmatrix}.$$

- **Market price of risk vector:**

$$\gamma = \begin{bmatrix} 2.7923 \\ 1.2016 \\ 1.7167 \end{bmatrix}.$$

- **Initial Factor Values:⁴**

$$X_0 = \begin{bmatrix} -0.0312 \\ 0.0384 \\ 0.0688 \end{bmatrix}.$$

These values are used as the initial condition for simulating the evolution of the yield curve factors over time.

3.10 Swaption Pricing Network

There is no closed-form solution for the swaption price provided by Equation (54), and thus the swaption sensitivity $\frac{\partial \text{PFS}_{t-1}}{\partial X_{t-1}^{(k)}}$ is not available in closed-form neither. To calculate such quantity we adopt a hybrid method that combines Monte Carlo simulation with neural network approximation.

Our goal is to approximate the mapping from the factor values and time to maturity to the corresponding swaption price. The approach proceeds as follows

- (1) **Simulate factor paths under the physical measure.** We simulate N_p sample paths of the factor process $(X_t)_{t=0,1,\dots,T_\beta}$ under the real-world (physical) measure $\mathbb{P}$. Since the factors

⁴Starting values X_0 correspond to filtered values of the factors process on January 31, 2022, which is the last of day of the sample that is used to estimate parameters. See Appendix B.2 of [41] for such values.

$X^{(1)}$, $X^{(2)}$, and $X^{(3)}$ are correlated, simulating full paths ensures that their interdependencies are preserved. This results in an array of $N_p \times (T_\beta + 1)$ samples.

- (2) **Select N_p samples of X_0 .** We sample the N_p initial values $\{(i)X_0\}_{i=1}^{N_p}$ from the simulated factor paths⁵ resulting from the previous step as the input dataset. These represent diverse and realistic starting conditions for the term structure factors.
- (3) **Randomize swaption maturities.** For each of the N_p samples, we generate an integer $T_\alpha^{(i)}$ uniformly at random from the set $\{1, 2, \dots, T_\alpha\}$. This assigns a random maturity to each sample of the previous step, reflecting a distribution of swaptions across different maturities.
- (4) **Compute swaption prices using the forward measure.** For each pair $((i)X_0, T_\alpha^{(i)})$, we use the forward-measure swaption pricing formula from [55]:

$$PS[0, \mathcal{T}^{(i)}, K, N] = \mathbb{E}^{\mathbb{Q}^{T_\alpha^{(i)}}} \left[\mathcal{P}_{T_\alpha^{(i)}} \mid \mathcal{F}_0 \right],$$

where $\mathcal{P}_{T_\alpha^{(i)}}$ denotes the discounted swap payoff at maturity $T_\alpha^{(i)}$, and $\mathcal{T}^{(i)}$ is the associated tenor structure. The expectation is evaluated using Monte Carlo simulation under the corresponding $T_\alpha^{(i)}$ -forward measure.

This process yields a training dataset consisting of input-output pairs $\{((i)X_0, T_\alpha^{(i)}), PS^{(i)}\}$, which is then used to train a neural network that approximates the pricing function, as described in Section 3.10.1, and its sensitivities for out-of-sample prediction. For the sensitivities, a finite difference approach is considered.

Once the swaption prices are obtained, we use the collection of factors, time to maturities and prices to train a fully connected Kolmogorov–Arnold neural network (KAN). The objective is to learn a smooth approximation of the pricing function, thereby enabling fast and accurate interpolation for out-of-sample factor inputs during hedging and simulation. This model is then

⁵Any value of the path, including non-initial values, can be selected.

used to evaluate swaption prices and sensitivities in real-time, significantly reducing computational overhead while preserving accuracy. The structure of the KAN network is discussed next.

3.10.1 Kolmogorov–Arnold Networks

Kolmogorov–Arnold Networks (KANs) are a recent neural architecture inspired by the Kolmogorov–Arnold representation theorem, which asserts that any multivariate continuous function can be written as a finite composition of univariate functions and addition. First introduced in [85], KANs provide a more interpretable and mathematically grounded alternative to standard FCNNs.

While traditional FCNNs apply fixed linear transformations followed by nonlinear activation functions (such as ReLU or Sigmoid), KANs take a different approach: they apply learnable univariate functions directly to each input coordinate and then combine the results linearly. This design allows KANs to adaptively learn nonlinear transformations on a per-dimension basis, making them highly expressive and naturally suited for capturing structured dependencies.

In this work, we use a KAN to approximate the swaption pricing function. Each univariate function is parameterized as a linear combination of Radial Basis Functions (RBFs), following the FastKAN framework proposed in [82]. These Gaussian RBFs are centered along the input axis, allowing the network to capture local patterns in the data with smooth transitions. This architecture enhances both flexibility and interpretability.

The structure of the KAN model used in this study is summarized as follows:

- Each input pair (X_τ, τ) , consisting of the yield curve factors at time $\tau \in \{0, 1, \dots, T_\alpha\}$ and the time-to-maturity, passes through a learnable univariate function composed of RBF kernels.
- The outputs of these functions are then linearly combined to produce the network output, which represents the swaption price under predefined parameters.
- Several such layers are stacked to form a deep KAN model, optionally including residual connections or normalization layers to improve training dynamics.

This structure provides the agent with a faster, more accurate, and interpretable approximation of swaption prices, which is particularly advantageous in settings where no closed-form pricing formula is available. For a detailed comparison highlighting why KAN are preferable to standard FCNN architectures in our setting, see Section 3.10.2.

Pricing Network Architecture. To approximate the swaption pricing function, we employ a three-layer Kolmogorov–Arnold Network (KAN) with layer widths of 8, 16, and 8, respectively. The model takes as input four features: the three term structure factors generated by the DTAFNS model and the time-to-maturity, expressed in years. The network outputs a single scalar corresponding to the price of a payer swaption.

Training data is generated using Monte Carlo simulations, with data for each swaption of maturity $T_\alpha^{(i)}$ being generated under the $T_\alpha^{(i)}$ -forward risk-neutral measure, using $N_p = 100,000$ sample paths. However, each swaption in the dataset has the same tenor and strike, each corresponding to these of the swaption we want to price in subsequent simulation experiments. The network is trained for up to 5,000 epochs using the Adam optimizer with a learning rate of 0.001, minimizing the smooth L1 loss (Huber loss) between predicted and simulated swaption prices.

Once trained, the pricing network is used to evaluate swaption prices and compute their gradients with respect to the input factors. We employ the central difference method to compute the gradient, as it exhibits greater stability and accuracy compared to the neural network’s gradient obtained through automatic differentiation. The computed sensitivities are then used to determine the hedge ratios in the benchmark rho-hedging strategies.

3.10.2 KAN vs. MLP Performance Comparison

This section presents a comparative analysis of the Kolmogorov–Arnold Network (KAN) and a standard FCNN for approximating swaption prices. While the KAN architecture consists of 3 layers with widths $[8, 16, 8]$, the FCNN model employs a deeper and wider architecture with 4 layers of respective widths $[512, 1024, 1024, 512]$. Despite its significantly smaller size, the

KAN demonstrates superior performance. Indeed, Figure 3.4 illustrates the in-sample (IS) and out-of-sample (OOS) performance trajectories of both models over 1000 training epochs.

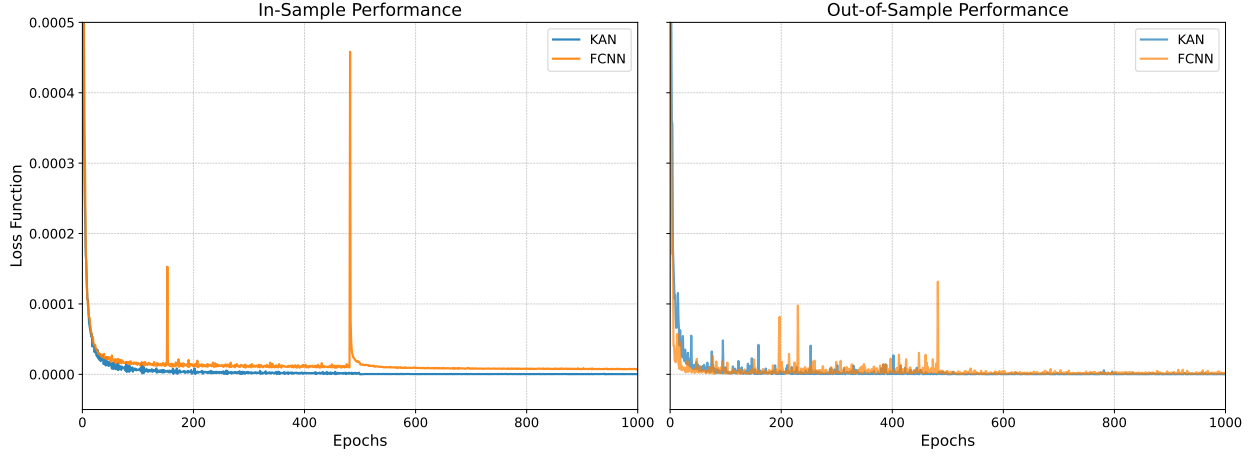


Figure 3.4: In-sample (left) and out-of-sample (right) performance of KAN and FCNN across 1000 training epochs.

The KAN consistently outperforms the MLP across multiple quantitative metrics. In the in-sample setting, the KAN exhibits a smooth and stable convergence trajectory with minimal variance. In contrast, the FCNN displays several sharp fluctuations, notably around epochs 150 and 500, indicating optimization instability and sensitivity to local minima.

Out-of-sample performance further highlights the KAN’s superiority in terms of generalization. The mean OOS error for the KAN is 1.73×10^{-6} . It is substantially lower than that of the FCNN, which is 4.71×10^{-6} . Additionally, the KAN achieves a significantly smaller generalization gap and reaches stable performance approximately 250 epochs earlier than FCNN (309 vs. 562 epochs). The final OOS error for the KAN remains in the order of 10^{-9} , an order of magnitude lower than that of the FCNN.

Overall, the KAN achieves:

- Lower mean and final IS and OOS errors,
- Smaller standard deviations, indicating greater robustness,
- Faster convergence to a stable solution,

- A tighter generalization gap, confirming enhanced predictive performance.

These findings support the conclusion that KAN offers a faster, more stable, accurate, and generalizable architecture for swaption price approximation compared to conventional feedforward networks, even when using significantly fewer parameters.

Conclusion

In this thesis, we present three research projects on stochastic processes and reinforcement learning, with applications in financial problems such as risk management, portfolio optimization, and hedging.

In the first study, we derive expressions for the potential densities of refracted skew Brownian motion (RSBM). Our main result is the determination of the general form of the probability density function of this process. As applications, we analyze its long-term asymptotic behavior. We also revisit earlier studies and show how our results provide a more general framework that encompasses them. Another contribution of this study is the proposal of two approaches for generating quasi-random samples by approximating the cumulative distribution function when exact random number generation is intractable. As a further application in risk management, we approximate Value at Risk (VaR) and Conditional Value at Risk (CVaR) using these methods.

The second project introduces a dynamic portfolio optimization model that accounts for both financial risk and greenhouse gas emissions associated with the portfolio. The model solves for the optimal mean-variance-GHG trade-off using reinforcement learning, enabling investors to achieve a balance between financial returns and environmental impact.

As a case study, we explore the optimal asset allocation across the 11 sectors of the S&P 500, customizing the analysis for investors with different levels of risk and GHG aversion. The findings show that optimal dynamic RL portfolios substantially outperform myopic dynamic, static and equal-weighted benchmarks. This superior performance is a consequence of (i) accurately

identifying sectors that strike the best balance between financial returns and environmental impact, in line with the investor's preferences, (ii) dynamically adjusting the portfolio in response to shifting market conditions, and (iii) integrating the impact of interaction between rebalancing decisions in the various stages of the investment horizon. The outperformance of the RL optimization model over the benchmarks is observed throughout the entire investment period, not just at the final time horizon, which highlights the robustness of the strategy. Moreover, this study provides investors with a valuable tool to assess the cost of altering their GHG emission levels or adjusting their risk constraints.

In the last essay, we set out to investigate whether reinforcement learning (RL) can provide a more effective and robust framework dynamic hedges of swaption positions compared to traditional sensitivity-based rho-hedging. Our approach incorporates multiple possible objective functions, namely mean squared error (MSE), downside risk, and CVaR. We examine how such different risk preferences translate into distinct hedging styles. We further explore the role of multiple hedging instruments, the impact of transaction costs, and the robustness of strategies under model misspecification.

The analysis yields several key insights. First, based on our three-factor arbitrage-free dynamic Nelson-Siegel model setting, we find that the incremental benefit of adding a second swap is substantial, while further improvements from a third instrument are modest. Within this structure, the first swap—the underlying of the swaption—naturally anchors the hedge by replicating the primary exposure to the level factor, while the second swap complements it by absorbing residual yield curve exposures or, under CVaR objectives, serving as a targeted tail-risk hedge.

Second, different objective functions lead to distinct hedging styles. The MSE agent leans heavily on level exposures, often gaining from directional risk but remaining more exposed to tail events. The downside risk agent adopts a more balanced posture, locking in gains when ahead and selectively taking risks when behind. The CVaR agent prioritizes tail protection most strongly, cutting exposures when in-the-money while dynamically seeking recovery premia when out-of-the-money. This adaptability across payoff scenarios demonstrates how RL policies can balance hedge

error reduction with selective premia harvesting, leading to more stable performance.

Third, our results highlight the structural advantages of reinforcement learning over rho-hedging. RL agents do not “over-hedge,” but instead optimally retain systematic exposures to risk premia, providing a more efficient risk–return tradeoff. They implicitly learn to reduce turnover in the presence of transaction costs, achieving lower implementation frictions than rigid sensitivity-based hedges. Moreover, RL strategies internalize pathwise dynamics and prove more robust under model misspecification. In particular, they demonstrate resilience against shocks on mean-reversion parameters, on top of outperforming rho-hedges in terms of controlling tail risks even under shocks on long-term factor levels. Taken together, these findings suggest that reinforcement learning offers a compelling framework for dynamic swaption hedging. By combining flexibility, adaptability and robustness, RL-based hedging strategies provide a meaningful improvement over traditional approaches in terms of risk management.

Bibliography

- [1] Guido Abate, Ignazio Basile, and Pierpaolo Ferrari. The integration of environmental, social and governance criteria in portfolio optimization: An empirical analysis. *Corporate Social Responsibility and Environmental Management*, 31(3):2054–2065, 2024.
- [2] Fabio Alessandrini and Eric Jondeau. Optimal strategies for ESG portfolios. *Swiss Finance Institute Research Paper*, (20-21), 2020.
- [3] Ed Alexander and D Poularikas. The handbook of formulas and tables for signal processing. *Boca Raton, Florida*, 33431, 1998.
- [4] GSI Alliance. Global sustainable investments review 2022. Technical report, GLOBAL SUSTAINABLE INVESTMENTS Alliance, 2022. URL <https://www.gsi-alliance.org/wp-content/uploads/2023/12/GSIA-Report-2022.pdf>.
- [5] Julian Amon, Margarethe Rammerstorfer, and Karl Weinmayer. Passive ESG portfolio management—the benchmark strategy for socially responsible investors. *Sustainability*, 13(16):9388, 2021.
- [6] Yunbi An and Wulin Suo. The compatibility of one-factor market models in caps and swaptions markets: Evidence from their dynamic hedging performance. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 28(2):109–130, 2008.
- [7] Théophile Anquetin, Guillaume Coqueret, Bertrand Tavin, and Lou Welgryn. Scopes of

- carbon emissions and their impact on green portfolios. *Economic modelling*, 115:105951, 2022.
- [8] Thilanka Appuhamillage, Vrushali Bokil, Enrique Thomann, Edward Waymire, and Brian Wood. Occupation and local times for skew brownian motion with applications to dispersion across an interface. 2011.
- [9] David Ardia, Keven Bluteau, Kris Boudt, and Koen Inghelbrecht. Climate change concerns and the performance of green vs. brown stocks. *Management Science*, 69(12):7607–7632, 2023.
- [10] Giacomo Ascione, Michele Bufalo, and Giuseppe Orlando. Modeling volatility of disaster-affected populations: A non-homogeneous geometric-skew brownian motion approach. *Communications in Nonlinear Science and Numerical Simulation*, 130:107761, 2024.
- [11] Yizhou Bai and Zhiyu Guo. An empirical investigation to the “skew” phenomenon in stock index markets: evidence from the nikkei 225 and others. *Sustainability*, 11(24):7219, 2019.
- [12] Yizhou Bai, Yongjin Wang, Haoyan Zhang, and Xiaoyang Zhuo. Bayesian estimation of the skew ornstein-uhlenbeck process. *Computational Economics*, pages 1–49, 2022.
- [13] Enrique Ballesterro, Mila Bravo, Blanca Pérez-Gladish, Mar Arenas-Parra, and David Plasamaria. Socially responsible investment: A multicriteria approach to portfolio selection combining ethical and financial objectives. *European Journal of Operational Research*, 216(2):487–494, 2012.
- [14] Olivier Bardou and Miguel Martinez. Statistical estimation for reflected skew processes. *Statistical inference for stochastic processes*, 13:231–248, 2010.
- [15] Sergey Berezin and Oleg Zayats. Skew brownian motion with dry friction: The pugachev-sveshnikov approach. *arXiv preprint arXiv:1705.10980*, 2017.

- [16] Florian Berg, Julian F Koelbel, and Roberto Rigobon. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6):1315–1344, 2022.
- [17] Thomas C Berry and Joan C Junkus. Socially responsible investing: An investor perspective. *Journal of Business Ethics*, 112:707–720, 2013.
- [18] BIS. OTC derivatives statistics at end-june 2024. Statistical release, Bank for International Settlements, 2024. URL https://www.bis.org/publ/otc_hy2411.htm. Accessed: 2025-06-26.
- [19] Fischer Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3(1-2): 167–179, 1976.
- [20] AN Borodin. Distributions of functionals of a skew brownian motion with discontinuous drift. *Journal of Mathematical Sciences*, 273(5):676–686, 2023.
- [21] Andrei N Borodin and Paavo Salminen. *Handbook of Brownian motion-facts and formulae*. Springer Science & Business Media, 2015.
- [22] Alan Brace, Dariusz Gatarek, and Marek Musiela. The market model of interest rate dynamics. *Mathematical Finance*, 7(2):127–155, 1997.
- [23] Alan Brace, Tim Dun, and Geoff Barton. Towards a central interest rate model. *Option Pricing, Interest Rates and Risk Management*, pages 278–313, 2001.
- [24] Damiano Brigo, Fabio Mercurio, et al. *Interest rate models: theory and practice*, volume 2. Springer, 2001.
- [25] Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.
- [26] Jay Cao, Jacky Chen, John Hull, and Zissis Poulos. Deep hedging of derivatives using reinforcement learning. *arXiv preprint arXiv:2103.16409*, 2021.

- [27] Li Chen, Lipei Zhang, Jun Huang, Helu Xiao, and Zhongbao Zhou. Social responsibility portfolio optimization incorporating esg criteria. *Journal of Management Science and Engineering*, 6(1):75–85, 2021.
- [28] Zengjing Chen, Panyu Wu, and Xiaowen Zhou. Optimal state equation for the control of a diffusion with two distinct dynamics. *arXiv preprint arXiv:2404.07618*, 2024.
- [29] Aleksander S Cherny, Albert N Shiryaev, and Marc Yor. Limit behavior of the” horizontal-vertical” random walk and some extensions of the donsker-prokhorov invariance principle. *Theory of Probability & Its Applications*, 47(3):377–394, 2003.
- [30] Jens HE Christensen, Francis X Diebold, and Glenn D Rudebusch. The affine arbitrage-free class of Nelson–Siegel term structure models. *Journal of Econometrics*, 164(1):4–20, 2011.
- [31] Guillaume Coqueret. Perspectives in ESG equity investing. *Perspectives in ESG equity investing (March 10, 2021)*, 2021.
- [32] TRA Corns and SE Satchell. Skew brownian motion and pricing european options. *The European Journal of Finance*, 13(6):523–544, 2007.
- [33] Marcel Andre Cresnik. Analysis of the hedging performance of select term structure models in a real world evolution framework. Master’s thesis, Universität Münster, 2015.
- [34] Marc Decamps, Marc Goovaerts, and Wim Schoutens. Self exciting threshold interest rates models. *International Journal of Theoretical and Applied Finance*, 9(07):1093–1122, 2006.
- [35] Francis X Diebold and Canlin Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2):337–364, 2006.
- [36] Gregor Dorfleitner and Mai Nguyen. A new approach for optimizing responsible investments dependently on the initial wealth. *Journal of Asset Management*, 18:81–98, 2017.
- [37] Gregor Dorfleitner and Sebastian Utz. Safety first portfolio choice based on financial and sustainability returns. *European Journal of Operational Research*, 221(1):155–164, 2012.

- [38] Gregor Dorfleitner, Michaela Leidl, and Johannes Reeder. Theory of social returns in portfolio choice with application to microfinance. *Journal of Asset Management*, 13:384–400, 2012.
- [39] Joost Driessen, Pieter Klaassen, and Bertrand Melenberg. The performance of multi-factor term structure models for pricing and hedging caps and swaptions. *Journal of Financial and Quantitative Analysis*, 38(3):635–672, 2003.
- [40] Darrell Duffie and Kenneth J Singleton. Modeling term structures of defaultable bonds. *The review of financial studies*, 12(4):687–720, 1999.
- [41] Ramin Eghbalzadeh, Patrice Gaillardetz, and Frédéric Godin. Evaluation of participating endowment life insurance policies in a stochastic environment. *European Actuarial Journal*, 14(2):467–494, 2024.
- [42] Ramin Eghbalzadeh, Frédéric Godin, and Patrice Gaillardetz. The discrete-time arbitrage-free Nelson-Siegel model: A closed-form solution and applications to mixed funds representation. *Annals of Actuarial Science*, 18(2):310–341, 2024.
- [43] Robert F Engle and Kevin Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, 2001.
- [44] Carlos Esparcia, Antonio Diaz, and Daniel Alonso. How important is green awareness in energy investment decisions? An environmentally-based rebalancing portfolio study. *Energy Economics*, 128:107174, 2023.
- [45] Rong Fan, Anurag Gupta, and Peter Ritchken. Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets. *The Journal of Finance*, 58(5):2219–2248, 2003.
- [46] Rong Fan, Anurag Gupta, and Peter Ritchken. On pricing and hedging in the swaption market: How many factors, really? *Journal of Derivatives*, 15(1):9, 2007.

- [47] Mark Freidlin and Shuenn-Jyi Sheu. Diffusion processes on graphs: stochastic differential equations, large deviation principle. *Probability theory and related fields*, 116:181–220, 2000.
- [48] Alexander Gairat and Vadim Shcherbakov. Density of skew brownian motion and its functionals with application in finance. *Mathematical Finance*, 27(4):1069–1088, 2017.
- [49] Alexander Gairat and Vadim Shcherbakov. Skew brownian motion with dry friction: Joint density approach. *Statistics & Probability Letters*, 187:109511, 2022.
- [50] Zhongqin Gao, Xiaowen Zhou, and Yan Lv. An optimal dividend problem for skew brownian motion with two-valued drift. *arXiv preprint arXiv:2402.11471*, 2024.
- [51] Stephan M Gasser, Margarethe Rammerstorfer, and Karl Weinmayer. Markowitz revisited: Social portfolio engineering. *European Journal of Operational Research*, 258(3):1181–1190, 2017.
- [52] Christopher C Geczy, John B Guerard Jr, and Mikhail Samonov. Warning: SRI need not kill your sharpe and information ratios-forecasting of earnings and efficient SRI and ESG portfolios. *The Journal of Investing*, 2020.
- [53] Lachlan Gibson, Marcus Hoerger, and Dirk Kroese. A flow-based generative model for rare-event simulation. *arXiv preprint arXiv:2305.07863*, 2023.
- [54] Alexandru Giurca and Svetlana Borovkova. Delta hedging of derivatives using deep reinforcement learning. *Available at SSRN 3847272*, 2021.
- [55] Frédéric Godin, Ramin Eghbalzadeh, and Patrice Gaillardetz. Pricing swaptions and zero-coupon futures options under the discrete-time arbitrage-free Nelson–Siegel model. *Review of Derivatives Research*, 26(2):171–206, 2023.
- [56] Anurag Gupta and Marti G Subrahmanyam. Pricing and hedging interest rate options: Evidence from cap–floor markets. *Journal of Banking & Finance*, 29(3):701–733, 2005.

- [57] Patrick S Hagan, Deep Kumar, Andrew S Lesniewski, and Diana E Woodward. Managing smile risk. *The Best of Wilmott*, 1:249–296, 2002.
- [58] John Michael Harrison and Lawrence A Shepp. On skew brownian motion. *The Annals of probability*, pages 309–313, 1981.
- [59] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1026–1034, 2015.
- [60] Markus Hirschberger, Ralph E Steuer, Sebastian Utz, Maximilian Wimmer, and Yue Qi. Computing the nondominated surface in tri-criterion portfolio selection. *Operations Research*, 61(1):169–183, 2013.
- [61] Thomas SY Ho and Sang-Bin Lee. Term structure movements and pricing interest rate contingent claims. *The Journal of Finance*, 41(5):1011–1029, 1986.
- [62] Marius Hofert, Avinash Prasad, and Mu Zhu. Quasi-random sampling for multivariate distributions via generative neural networks. *Journal of Computational and Graphical Statistics*, 30(3):647–670, 2021.
- [63] Felix Horger, Tobias Würfl, Vincent Christlein, and A Maier. Deep learning for sampling from arbitrary probability distributions. *arXiv preprint arXiv:1801.04211*, 2018.
- [64] John Hull and Alan White. Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3(4):573–592, 1990.
- [65] IPCC. Climate change 2022: Mitigation of climate change. Technical Report Sixth Assessment Report, Intergovernmental Panel on Climate Change (IPCC), 2022. URL <https://www.ipcc.ch>.
- [66] Kiyosi Itô, P Henry Jr, et al. *Diffusion processes and their sample paths*. Springer Science & Business Media, 2012.

- [67] Sebastian Jaimungal, Silvana M Pesenti, Ye Sheng Wang, and Hariom Tatsat. Robust risk-aware reinforcement learning. *SIAM Journal on Financial Mathematics*, 13(1):213–226, 2022.
- [68] Farshid Jamshidian. LIBOR and swap market models and measures. *Finance and Stochastics*, 1:293–330, 1997.
- [69] Robert A Jarrow, David Lando, and Stuart M Turnbull. A markov model for the term structure of credit risk spreads. *The review of financial studies*, 10(2):481–523, 1997.
- [70] Eric Jondeau and Michael Rockinger. Conditional asset allocation under non-normality: how costly is the mean-variance criterion? *Available at SSRN 674424*, 2005.
- [71] Ioannis Karatzas and Steven E Shreve. Trivariate density of brownian motion, its local and occupation times, with application to stochastic control. *The Annals of Probability*, pages 819–828, 1984.
- [72] Samuel Karlin. *A first course in stochastic processes*. Academic press, 2014.
- [73] Alexander Kempf and Peer Osthoff. The effect of socially responsible investing on portfolio performance. *European financial management*, 13(5):908–922, 2007.
- [74] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [75] Dieter Kraft. A software package for sequential quadratic programming. *Forschungsbericht-Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt*, 1988.
- [76] Davide Lauria, W Brent Lindquist, Stefan Mittnik, and Svetlozar T Rachev. ESG-valued portfolio optimization and dynamic asset pricing. arxiv. *arXiv preprint arXiv:2206.02854*, 2022.

- [77] J.-F Le Gall. One-dimensional stochastic equations involving the local times of the unknown process. *Proceedings of the International Conference on Stochastic Analysis and Applications, Swansea, 11-15 April 1983, Lecture Notes in Mathematics*, 1095:51–82, 1984.
- [78] J.-F Le Gall. One-dimensional stochastic differential equations involving the local times of the unknown process. *Stochastic analysis and applications*, 1985.
- [79] Théo Le Guenedal, François Lombard, Thierry Roncalli, and Takaya Sekine. Net zero carbon metrics. *Available at SSRN 4033686*, 2022.
- [80] Antoine Lejay. On the constructions of the skew brownian motion. 2006.
- [81] Antoine Lejay and Sara Mazzonetto. Maximum likelihood estimator for skew brownian motion: the convergence rate. *Scandinavian Journal of Statistics*, 2023.
- [82] Ziyao Li. Kolmogorov-Arnold networks are radial basis function networks. 2024.
- [83] Konstantinos Liagkouras, K Metaxiotis, and G Tsihrintzis. Incorporating environmental and social considerations into the portfolio optimization process. *Annals of Operations Research*, pages 1–26, 2022.
- [84] Yong-Jun Liu, Guo-Sen Yang, and Wei-Guo Zhang. A novel regret-rejoice cross-efficiency approach for energy stock portfolio optimization. *Omega*, 126:103051, 2024.
- [85] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljačić, Thomas Y Hou, and Max Tegmark. KAN: Kolmogorov-Arnold networks. *arXiv preprint arXiv:2404.19756*, 2024.
- [86] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.
- [87] Chari Maree and Christian W Omlin. Balancing profit, risk, and sustainability for portfolio management. In *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, pages 1–8. IEEE, 2022.

- [88] Andrea Mazzon and Peter Tankov. Optimal stopping and divestment timing under scenario ambiguity and learning. *arXiv preprint arXiv:2408.09349*, 2024.
- [89] Oskari Mikkilä and Juho Kanninen. Empirical deep hedging. *Quantitative Finance*, 23(1): 111–122, 2023.
- [90] Diganta Misra. Mish: A self regularized non-monotonic activation function. *arXiv preprint arXiv:1908.08681*, 2019.
- [91] Sanjay K Nawalkha, Natalia Beliaeva, and Gloria M Soto. A new taxonomy of the dynamic term structure models. *Available at SSRN 1265286*, 2010.
- [92] Charles R Nelson and Andrew F Siegel. Parsimonious modeling of yield curves. *Journal of Business*, pages 473–489, 1987.
- [93] Daniel B Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the econometric society*, pages 347–370, 1991.
- [94] Kenjiro Oya. Deep hedging Bermudan swaptions. *arXiv preprint arXiv:2411.10079*, 2024.
- [95] L’uboš Pástor, Robert F Stambaugh, and Lucian A Taylor. Sustainable investing in equilibrium. *Journal of financial economics*, 142(2):550–571, 2021.
- [96] Lasse Heje Pedersen, Shaun Fitzgibbons, and Lukasz Pomorski. Responsible investing: The ESG-efficient frontier. *Journal of financial economics*, 142(2):572–597, 2021.
- [97] Reilly Pickard and Yuri Lawryshyn. Deep reinforcement learning for dynamic stock option hedging: A review. *Mathematics*, 11(24):4943, 2023.
- [98] Robert Piessens, Elise de Doncker-Kapenga, Christoph W Überhuber, and David K Kahaner. *Quadpack: a subroutine package for automatic integration*, volume 1. Springer Science & Business Media, 2012.

- [99] PRI. Principales for responsible investments annual report 2018. Technical report, Principales for Responsible Investments, 2018. URL <https://www.unpri.org/annual-report-2018>.
- [100] Jorge Ramirez. Multi-skewed brownian motion and diffusion in layered media. *Proceedings of the American Mathematical Society*, 139(10):3739–3752, 2011.
- [101] R Tyrrell Rockafellar, Stanislav Uryasev, et al. Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2:21–42, 2000.
- [102] Erik Štrumbelj and Igor Kononenko. Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41(3):647–665, 2014.
- [103] Tim Dun, Tim Dun, G.W. Barton, Geoff Barton, Erik Schlögl, and Erik Schlogl. Simulated swaption delta-hedging in the lognormal libor model. *International Journal of Theoretical and Applied Finance*, 4(4):677–709, 2001.
- [104] UNFCCC. Paris agreement. United Nations Framework Convention on Climate Change (UNFCCC), December 2015. Available online: https://unfccc.int/sites/default/files/english_paris_agreement.pdf.
- [105] U.S.Government. The long-term strategy of the united states: Pathways to net-zero greenhouse gas emissions by 2050. United States Government, November 2021. Available online: <https://www.whitehouse.gov/wp-content/uploads/2021/10/US-Long-Term-Strategy.pdf>.
- [106] Sebastian Utz, Maximilian Wimmer, Markus Hirschberger, and Ralph E Steuer. Tri-criterion inverse portfolio optimization with application to socially responsible mutual funds. *European Journal of Operational Research*, 234(2):491–498, 2014.
- [107] Pieter M van Staden, Peter A Forsyth, and Yuying Li. A parsimonious neural network

- approach to solve portfolio optimization problems without using dynamic programming. *arXiv preprint arXiv:2303.08968*, 2023.
- [108] Oldrich Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.
- [109] Nhi NY Vo, Xuezhong He, Shaowu Liu, and Guandong Xu. Deep learning for decision making and the optimization of socially responsible investments and portfolio. *Decision Support Systems*, 124:113097, 2019.
- [110] John B Walsh. A diffusion with a discontinuous local time. *Astérisque*, 52(53):37–45, 1978.
- [111] Haojie Wang, Han Chen, Agus Sudjianto, Richard Liu, and Qi Shen. Deep learning-based BSDE solver for LIBOR market model with application to Bermudan swaption pricing and hedging. *arXiv preprint arXiv:1807.06622*, 2018.
- [112] Suxin Wang, Shiyu Song, and Yongjin Wang. Skew ornstein–uhlenbeck processes and their financial applications. *Journal of Computational and Applied Mathematics*, 273:363–382, 2015.
- [113] Luluk Widyawati. A systematic literature review of socially responsible investment and environmental social governance metrics. *Business Strategy and the Environment*, 29(2): 619–637, 2020.
- [114] Qun Wu, Xinwang Liu, Jindong Qin, Ligang Zhou, Abbas Mardani, and Muhammet Deveci. An integrated multi-criteria decision-making and multi-objective optimization model for socially responsible portfolio selection. *Technological Forecasting and Social Change*, 184: 121977, 2022.
- [115] Panos Xidonas and Eric Essner. On ESG portfolio construction: A multi-objective optimization approach. *Computational Economics*, 63(1):21–45, 2024.

- [116] Ming Zhang. Calculation of diffusive shock acceleration of charged particles by skew brownian motion. *The Astrophysical Journal*, 541(1):428, 2000.
- [117] Song-Ping Zhu and Xin-Jiang He. A new closed-form formula for pricing european options under a skew brownian motion. *The European Journal of Finance*, 24(12):1063–1074, 2018.