Deep hedging of CVA

Oscar Gummesson Atroshi

Osman Sibai

LUND UNIVERSITY

Department of Automatic Control

MSc Thesis
TFRT-6253
ISSN 0280-5316

Department of Automatic Control
Lund University
Box 118
SE-221 00 LUND
Sweden

Printed in Sweden by Tryckeriet i E-huset
Lund 2024

Abstract

Large financial institutions are vulnerable to numerous financial risks, necessitating robust regulatory frameworks to prevent crises such as those experienced in 2008. The Basel framework, devised by the Basel Committee on Banking Supervision, incorporates critical measures such as the *credit valuation adjustment* (CVA) to mitigate these risks. CVA fluctuates significantly based on market factors and counterparty conditions, these fluctuations need to be handled, and this is done through hedging. Hedging CVA is challenging due to its sensitivity to dynamic market conditions and the complexity of underlying assets, compounded by factors such as cross-gamma and wrong-way risk, which add significant complexity to effective risk management. This study explores the use of deep hedging, employing reinforcement learning to devise robust hedging strategies that navigate the complexities often associated with traditional analytic models. Through experimental simulations, this research compares the efficacy of traditional delta hedging with that of a reinforcement learning-based strategy, providing insights into their respective performances. The study evaluates two different market models, with the RL strategies showing promising results, particularly in the less complex model, highlighting the challenges of addressing high-dimensional problems. The findings establish a foundation for further research and demonstrate the potential of reinforcement learning in enhancing CVA hedging strategies.

Acknowledgements

We extend our heartfelt gratitude to our supervisor from Nordea Markets, Shengyao Zhu, for providing us with invaluable insights into the practical aspects of CVA. Additionally, we are deeply appreciative of the guidance and expertise provided by our supervisors from LTH, Manu Upadhyaya and Magnus Wiktorsson, who have greatly aided our understanding of the theoretical framework and enhanced our skills in academic writing. Our thanks also go to the Department of Automatic Control for their unwavering support throughout this project. We are especially grateful to the team at StableBaselines3; their Reinforcement Learning Python code was instrumental in the completion of our research. Lastly, we wish to thank our examiner, Johan Eker, for dedicating his time to evaluate this thesis.

Contents

1

Introduction

1.1 Background and motivation

Banks manage large trading books with exposure to various financial risks. To avoid scenarios like the 2008 financial crisis, oversight institutions have introduced laws and regulations to manage these risks. One crucial set of regulations is the *Basel framework*, developed by the *Basel Committee on Banking Supervision*. Ensuring compliance with banking regulations, including the Basel framework, is a critical responsibility of financial institutions [Buehler et al., 2015].

Large banks have developed advanced in-house control systems to ensure compliance with regulations. Traditional financial models have been used to generate various risk exposure measures for risk management. However, recent research indicates significant potential in modern machine learning approaches [Buehler et al., 2018]. Many traditional methods fail to consider transaction costs and additional market information [Buehler et al., 2018]. With the technology known as *deep hedging*, machines can analyze large amounts of historical data to make more advantageous hedging decisions. Hedging is reducing financial risk by taking offsetting positions in different assets. Machine learning methods are implemented with a clear objective, meaning they are trained on a specific, explicit reward function. This is particularly useful for addressing risk management questions where one wants to explore different risk measures.

This thesis is done in collaboration with *Nordea Markets* in Copenhagen, part of *Nordea Bank*, a Nordic financial services group operating in northern Europe. In this thesis, we explore hedging in the specific context of a *credit value adjustment* (CVA). CVA originates from the *third Basel accords* [Bank for International Settlements, 2020] and has been a significant component of P&L performance for Nordea and, more broadly, any large global financial institution offering derivatives. P&L stands for *profit and loss*, which measures financial gains and losses over time. The cross-gamma effect of CVA refers to the nonlinear relationship between different risk factors, complicating the hedging process. Thus, developing a machine-learning approach to address this issue can help banks improve their risk management strategies by providing more accurate and nuanced insights.

1.2 Objectives

The goal of this thesis is to develop strategies to effectively manage the risks associated with CVA of derivatives using advanced machine learning techniques. Specifically, we explore the use of deep hedging—hedging strategies based on deep *reinforcement learning* (RL)—within simulated market environments. Our research objectives are

i) the development of deep hedging to manage the risks associated with CVA in simulated environments,

ii) the evaluation of deep hedging against a baseline method called *delta hedging*, and

iii) evaluation of the performance of the methods with respect to trading costs and underlying correlations in the simulated market environments.

1.3 Significance

This thesis addresses critical financial stability issues that became apparent during the 2007-2008 global financial crisis, during which banks incurred substantial losses due to CVA [Pandit, 2023]. Understanding and effectively managing CVA is essential because it represents the cost of potential counterparty defaults [Brigo et al., 2013]. Accurate CVA management is crucial for maintaining the integrity and stability of financial institutions, especially those dealing with over-the-counter derivatives.

CVA hedging is vital to modern risk management, offering protection against counterparty risk and helping institutions comply with regulatory requirements. However, the complexities in accurately quantifying CVA, market dynamics, model risk, and operational challenges make it difficult. Continuous advancements in risk management practices and technologies are necessary to address these challenges effectively [Gregory, 2012]. Additionally, complexities like *wrong-way risk* (WWR) and cross-gamma further complicate the development of effective risk management strategies [Hull and White, 2012; Chicot, 2019].

Research has shown that RL-based hedging strategies can outperform traditional methods, especially under conditions of transaction costs [Kolm and Ritter, 2019; Daluiso et al., 2023]. These advanced strategies leverage reinforcement learning to adapt to dynamic market conditions, providing a more robust framework for managing financial risks.

This thesis contributes to improving financial risk management practices by developing and validating RL-based strategies. It supports better preparedness and resilience against future market disruptions, making it highly relevant for financial institutions, regulators, and researchers aiming to enhance the stability and robustness of the global financial system.

1.4 Scope and methodology

This thesis focuses on developing and validating RL-based strategies for hedging CVA in derivative portfolios. The RL algorithms are implemented and tested to develop robust hedging strategies for CVA. The performance of these RL-based strategies is compared to traditional hedging methods, with particular attention to their effectiveness under different correlation conditions for the underlying market processes and the influence of transaction costs. The research is conducted within a simulated financial environment to evaluate the efficacy of RL agents in learning to hedge. This environment incorporates key factors such as WWR, cross-gamma, transaction costs, and correlations between underlying processes. The RL strategies are examined under these conditions to evaluate their performance. A limitation is that they are tested in the same environment, though on different data, as trained.

The methodology of the thesis includes several key steps to achieve these objectives. First, a comprehensive review of existing literature on traditional hedging methods and recent advancements in the use of RL for hedging is conducted.

Next, two simulated financial environments are developed to represent market dynamics and conditions for CVA. These environments are designed to accurately reflect the complexities of the financial markets, including WWR, cross-gamma, transaction costs, and the correlations between underlying processes. The correlations of the underlying assets are assumed to be constant over time, and the transaction costs are assumed to be linear. The first of these environments is extremely simple, with just two underlying assets, built on the assumption that the contract it is based on has only one payment very far into the future. This environment implements the simplest model of a CVA that still captures the non-linear complexities. The second one is based on a CVA for a real interest rate product. The CVA in this second environment is simulated according to models by [Brigo et al., 2013]. RL algorithms, specifically the *proximal policy optimization* (PPO) algorithm, are implemented in these simulated environments. The state spaces, action spaces, and reward functions are defined to align with the CVA hedging problem.

The RL agents are trained in these simulated environments, allowing them to learn optimal hedging strategies through interaction and optimization. It is assumed in both environments that there are no liquidity constraints and that trading can be done directly in the underlying assets. The performance of the RL-based strategies are evaluated by comparing them with the traditional delta hedging method, and the naive method of not hedging. Performance metrics relating to P&L are defined and used for these comparisons on numerical simulations. While we check whether the strategies make net profits or losses, we do not keep track of how much cash injection they require during the hedging process to take their positions, we focus instead on the variance and risk metrics. These simulations analyze the performance of RL-based strategies under various market conditions and correlation scenarios.

The results of the comparative analyses are presented, highlighting the strengths and weaknesses of RL-based strategies. The findings are discussed regarding their implications for financial applications and the broader effort to improve financial risk management practices. The thesis concludes by suggesting areas for further research, emphasizing the potential of reinforcement learning in developing effective and robust hedging strategies for CVA. This comprehensive approach aims to enhance stability and robustness in risk management practices.

1.5 Related work

The benchmark for hedging, when assuming no trading costs and continuous trading, is established by the minimum variance hedge, as described by Hull and White [Hull and White, 2017]. Furthermore, [Davis et al., 1993] developed a set of complex nonlinear *partial differential equations* (PDEs) that provide optimal solutions for continuous trading with trading costs in Black and Scholes markets. However, these PDEs become computationally unfeasible in dimensions higher than one.

Recent advancements in financial technology have seen numerous research groups pivot towards utilizing *reinforcement learning* (RL) for hedging applications. [Buehler et al., 2018] demonstrated that optimal hedging strategies applicable to a wide range of market conditions—including various market frictions like trading costs, market impact, and liquidity constraints—can theoretically be approximated well by standard reinforcement learning algorithms. They also provided practical evidence of their efficacy by successfully hedging call options in a simulated market environment.

Moreover, [Kolm and Ritter, 2019] illustrated that RL could effectively learn to hedge even without complete information, using a log-normal market as a case study where perfect hedging is unachievable. Following this, [Du et al., 2020] explored deeper into RL's capabilities by training models using deep Q-learning, pop-art, and *proximal policy optimization* (PPO). Their findings suggested that models utilizing PPO performed optimally and trained more rapidly than others, achieving superior results in hedging call options with discrete trading and associated trading costs.

[Cao et al., 2021] further confirmed RL's potential in deriving optimal hedging strategies for call options under trading costs, utilizing a Black and Scholes market framework and a stochastic volatility market model. Their approach involved an innovative Q-learning model that calculates two distinct Q-functions to effectively manage expected and squared costs.

Building on the work of [Du et al., 2020], our study expands the application of PPO-based RL models to manage more complex scenarios involving multiple underlying assets and intricate market models, aiming to achieve better performance over traditional delta hedging techniques. This progression underscores RL's broad applicability and robustness in financial hedging, especially for managing CVA, as preliminarily evidenced by promising results in [Daluiso et al., 2023].

1.6 Outline

In Chapter 2 of this thesis, we introduce the essential financial background needed to comprehend the subsequent discussions. We explore important elements of financial markets with a specific focus on credit risk management, highlighting the motivation behind the introduction of CVA into contemporary financial practices. This chapter sets the stage for understanding the complex interactions and dependencies that define the risk landscape in financial systems.

Moving into Chapter 3, we introduce the concept of hedging. We also delve into the mathematical underpinnings of stochastic control as a way to represent and solve the dynamic hedging problem. This section is crucial as it lays down the theoretical framework for our hedging strategies.

Chapter 4 transitions from theory to the practical application of these concepts, introducing the stochastic models used to simulate the financial markets in which our agents will operate. We discuss the assumptions, variables, and the construction of these models to reflect real-world market behaviors as closely as possible.

In Chapter 5, we take a deeper look at the specific contract we are focusing on — the CVA. This chapter outlines the unique challenges associated with hedging CVA, including its sensitivity to market factors and its critical role in managing counterparty risk.

Following that, Chapter 6 details the specific environments where our hedging strategies will be applied. It outlines the operational settings, the conditions under which our models will function, and how we will evaluate their performance. This chapter is designed to bridge the gap between theoretical models and their practical implementation.

Chapter 7 covers the theoretical background on reinforcement learning, the core algorithm behind our hedging models. We discuss the specific algorithms used, their suitability for our needs, and how they adapt to the unique challenges of financial hedging.

Beginning Chapter 8, we describe the setup for our simulations and models, detailing which variables we vary and the values they can take. This chapter then presents the results of our simulations, providing a critical analysis of the effectiveness of deep hedging strategies in managing CVA.

In Chapter 9, we discuss the implications of our findings and analyze what the results suggest about the potential and limitations of using deep learning algorithms for dynamic hedging.

Finally, Chapter 10 concludes the thesis by summarizing the broader implications of this work and discussing potential future directions for research in this area. We reflect on how the findings could influence future developments in financial modeling, particularly focusing on the integration of machine learning techniques in risk management strategies.

1.7 Individual Contributions

The authors have jointly contributed to the majority of the project, sharing similar backgrounds in financial modeling, machine learning, and programming, which facilitated a collaborative approach to understanding the problem. The writing process was iterative, with both authors actively contributing to each chapter, ensuring a balanced input throughout the development of the thesis.

1.8 Implementation code

For the interested reader, the code used is available at the GitHub repository: `https://github.com/ososib/Deep-Hedging-of-CVA.git`. The code contains the market simulators, RL implementation, evaluation scripts, and helper functions we used in our thesis.

2

Financial instruments and implications

There exists a wide array of financial instrument types. We care about some from a diverse subset of them known as derivatives. In this chapter, we will introduce financial derivatives and present some use cases and a few examples. We also introduce some framework and regulatory details governing the operation of financial derivatives. We will also discuss counterparty credit risk and explain briefly how it is managed.

2.1 Financial derivatives

Derivatives are financial instruments whose value depends on other instruments called underlying assets. Derivatives have existed and been traded for a very long time, there is evidence of them being used as far back as the ancient Greeks [Oosterlinck, 2017]. Today there is an incredibly massive and diverse derivatives market, with the notional value of outstanding OTC derivatives in June 2023 being $715 trillion [Bank for International Settlements, 2023]. They are used for various purposes, including hedging risk, speculating on price movements, and gaining access to otherwise inaccessible assets or markets. The most common types of financial derivatives are forwards, futures, options, and swaps [Hull, 2018]. In this thesis, we will focus on swaps and swaptions.

OTC derivatives

Over-the-counter (OTC) derivatives are bespoke financial contracts that are traded in decentralized markets, allowing for significant customization to meet the specific needs of the involved parties. These instruments are not traded through formal exchanges and thus are not subject to the same standardization as exchange-traded derivatives [Heckinger et al., 2014]. OTC derivatives encompass a variety of instruments, each tailored to the risk management, investment, or speculative needs of the counterparties [Hull, 2018].

While OTC derivatives offer the benefits of customization and potential cost savings due to the absence of exchange fees, they also introduce risks like counterparty credit risk, as there is no central clearinghouse to back the performance of these contracts [Heckinger et al., 2014]. The regulatory landscape for OTC derivatives has evolved significantly over time, particularly with reforms such as the Dodd-Frank Act's Titles VII and VIII, which aim to balance market vibrancy with risk mitigation [Tarbert, 2020].

Zero-coupon bonds

A fundamental example of an interest rate derivative is the *zero-coupon bond* (ZCB). This financial instrument illustrates the concept that future money holds less value than present money, necessitating a discounting method for projecting future gains and losses to the present value. The discounting factor used for this purpose is called the discount factor $D(t,T)$, which varies depending on the time points t (current time) and T (future time) as well as the prevailing interest rates.

A ZCB is a financial contract that promises to pay a fixed amount — typically one unit of currency — at a predetermined future date. The present value of a ZCB at time t for a payoff at time T (maturity time) is expressed as

$$P(t,T) = \mathbb{E}[D(t,T)].$$

Taking the expectation is necessary here because the discount factor is influenced by the interest rate, which is typically modeled as a stochastic process.

Due to their simplicity, zero-coupon bonds are not traded as standalone entities in actual markets. Instead, they serve as foundational elements in the development of pricing models for more complex interest rate-based derivatives [Björk, 2019].

Swaps

Swaps are contractual agreements between two entities to exchange cash flows or other financial instruments over a designated period. Among the most prevalent types of swaps are interest rate swaps, currency swaps, and commodity swaps.

Example. Consider two companies: Company A has a loan with a floating interest rate, and Company B has a loan with a fixed interest rate, both of equivalent notional value. Through an interest rate swap, Company A agrees to pay fixed interest rates to Company B, while receiving floating rates in return. This arrangement allows both companies to tailor their exposure to interest rate fluctuations in alignment with their financial strategies.

An important financial instrument for this thesis is the *interest rate swap* (IRS), which involves (as described above) the exchange of one type of interest rate payment for another, typically from fixed to floating or vice versa, between two parties.

Swaptions

A swaption is a derivative that provides the holder with the option to enter into a given swap agreement at a given future date. It offers flexibility to opt into a swap when favorable or to decline if the conditions are not advantageous.

Our focus is particularly on interest rate swaptions, which are options on interest rate swaps. The value of the underlying IRS is determined by the fixed rate K and the floating rate r_t. In contrast, the valuation of a swaption not only depends on these rates but also on the underlying swap rate, which is the rate that would make the swap's expected value zero today.

The fixed rate K should be set so that the expected value of the swap at the time of issuing the contract is zero, assuming no default risk from the counterparty. This value can be found by using ZCBs. The formula to calculate K is

$$K = \frac{P(0,T_a) - P(0,T_b)}{\sum_{i=a+1}^{b} \tilde{\beta}_i P(0,T_i)},$$

where $P(t,T)$ represents the price of the ZCBs at time t with maturity T, T_a is the start time of the swap, T_b is the end time of the swap, and $\tilde{\beta}_i$ are the interval lengths between consecutive payments, $T_{i+1} - T_i$.

Similarly, the underlying swap rate, necessary for the swaption pricing, is adjusted continuously and computed to ensure the swaption's expected value remains neutral. It is calculated as:

$$S_{s,T_b}(t) = \frac{P(t,s) - P(t,T_b)}{\sum_{i=a+1}^{b} \tilde{\beta}_i P(t,T_i)}$$

Here, $S_{s,T_b}(t)$ is the swap rate at time t between the start time s and the end time T_b. The rate is calculated using the same variables and principles as for K, but evaluated at different times to reflect changing market conditions.

2.2 Counterparty credit risk

Default is a critical aspect of risk modeling, though defining it precisely is complex. In 1999, the *International Swaps and Derivatives Association* (ISDA) identified six types of credit events that can be considered defaults in the context of financial derivatives [Brigo et al., 2013]. These credit events represent different ways of breaking the terms of a debt contract, with most involving a failure to pay.

For our purposes, we assume that no contract payments occur after default and that a fixed percentage of the owed amount is recovered. This assumption is common in practice because debts are often sold to other companies at a fixed percentage of their value to avoid lengthy court processes.

[Gregory, 2010] explains that counterparty credit risk is the risk associated with OTC derivative and security financing transactions, where one party may default before the contract expires. This risk is more significant in derivatives due to the larger market size and complexity compared to security financing transactions.

Counterparty credit risk presents unique challenges in derivatives transactions. For instance, exposure in these derivatives is highly volatile, which complicates risk management [Zhu and Pykhtin, 2007]. Several critical factors are used to evaluate this risk: the present value, which is the expected value of future cash flows discounted at the risk-free rate, crucial for valuation; exposure, or the maximum potential loss should the counterparty default, which is the positive present value of a contract; the probability of default, often derived from market data like credit ratings and quantified as a cumulative distribution function over time; the recovery rate, which measures the percentage of exposure recoverable in a default; and the loss-given-default, indicating the proportion of exposure that is lost if a default occurs [Brigo et al., 2013]. These factors together frame the complex nature of counterparty credit risk in financial derivatives.

2.3 Managing credit risk

Strategies to control and mitigate credit risk include centralized clearing houses, netting agreements, and collateralization [Brigo et al., 2013]. This discussion focuses on active hedging, typically managed by a bank's valuation adjustments team.

Traditionally, the pricing of derivatives has employed risk-neutral valuation methods to establish a fair price, often without taking into account the credit quality of the counterparty [Ahlberg, 2013]. [Gregory, 2010] discusses strategies for managing counterparty credit risk, such as setting risk limits and conducting transactions only with parties of sufficient credit quality. However, counterparty risk was seldom included in derivative pricing, leading to the development of credit valuation adjustments to account for trade riskiness. These adjustments fluctuate in value and require efficient management and hedging, which we will explore in detail in Section 5, including pricing models and hedging strategies.

2.4 Basel III

Basel III is a comprehensive set of reform measures designed to enhance the regulation, supervision, and risk management within the banking sector, with a focus on counterparty credit risk. Developed by the Basel Committee on Banking Supervision, Basel III was established in response to the deficiencies in financial regulation revealed by the financial crisis of 2007-09. The framework aims to fortify banks by increasing their ability to absorb shocks arising from financial and economic stress, whatever the source, thus improving risk management and governance as well as strengthening banks' transparency and disclosures [*Basel III: International Regulatory Framework for Banks* 2021]. Importantly for this study, it is the regulation that stipulates that banks must consider the credit risks associated with their derivatives in the form of CVA.

3

Hedging

In this chapter, we introduce the concept of hedging. We also present a formal mathematical model for the dynamic hedging problem.

3.1 Hedging

Hedging is a strategy designed to mitigate risk by adopting positions negatively correlated with current holdings, known as the hedge. The primary objective is to balance potential losses in one area with gains in another. This approach inherently involves a risk-reward trade-off: while hedging can limit potential losses, it may also cap potential gains and incurs costs associated with acquiring the hedging instruments.

Example. Company X, based in Sweden, exports machinery parts to the United States and anticipates receiving a payment of $1 million in six months. This payment is subject to foreign exchange risk due to potential fluctuations in the exchange rate between the dollar and the Swedish krona (SEK). To hedge against this risk, Company X enters into a forward contract with a bank to exchange $1 million for SEK at a predetermined exchange rate on the payment date.

- If the dollar depreciates against the SEK, Company X will still receive the fixed amount in SEK as stipulated in the forward contract, thus mitigating any loss that would have resulted from the exchange rate movement.
- Conversely, if the dollar appreciates against the SEK, Company X will not benefit from the stronger dollar as they are locked into the agreed rate, missing out on potential additional earnings.

This forward contract ensures that Company X's revenue from the export remains predictable regardless of currency volatility, although it does restrict the company from benefiting from favorable rate movements.

3.2 Hedging as a stochastic optimal control problem

The hedging problem can be rephrased as a finite-horizon discrete-time stochastic optimal control problem. We consider a discrete-time financial market with finite time horizon T, trading times $0 = t_0 < t_1 < \ldots < t_N = T$, and stochastic final index $\tilde{N}$, such that the final time is $\min(t_N, t_{\tilde{N}})$, where t_N is the default-free final time and $t_{\tilde{N}}$ is the time when default occurs if it occurs. Say there are n hedging instruments and one liability to hedge. We define $p = \{p_{t_i}\}_{i=0}^N$ as an $\mathbb{R}^n$-valued stochastic process where $[p_{t_j}]_i$ represents the price of asset $i \in \{1, \ldots, n\}$ at time t_j, and $\{l_{t_i}\}_{i=0}^N$ be an $\mathbb{R}$-valued stochastic process where l_{t_j} represents the value of the liability at time t_j.

To hedge the liability, we may trade in the assets in p. We define the $\mathbb{R}^n$-valued stochastic process $q = \{q_{t_i}\}_{i=0}^N$ such that $[q_{t_j}]_i$ represents the amount of asset i owned at time t_j, and purchased at time t_{j-1}. For each $j \in \{1, \ldots, N\}$, q_{t_j} can only be constructed using information up until time t_{j-1}. We also say we begin with no holdings, that is $q_{t_0} = 0$. Realistically there should be limits to such trading strategies due to liquidity and trading restrictions, but we will not model those. Our only limit is that $[q_{t_j}]_i$ should be finite for all assets i and times t_j. We call the set of such admissible trading strategies $\mathscr{Q}$.

The goal of the hedge is for changes in the value of the hedge to track changes in the value of the liability. That is, we want to find strategies that pick q_{t_i} according to

$$\underset{q_{t_i}}{\text{minimize}}\ \mathbb{E}\left[\mathscr{R}\big((l_{t_i} - l_{t_{i-1}}) - (p_{t_i} - p_{t_{i-1}})^\top q_{t_i}\big)\middle| l_{t_{i-1}}, p_{t_{i-1}}\right], \tag{3.1}$$

for some risk function $\mathscr{R} : \mathbb{R} \to \mathbb{R}$ with a minimum at 0. Moreover, we would like to account for transaction costs. In literature, there are three ways this can be done. There can be a fixed cost associated with making a trade as in [Muhle-Karbe et al., 2017], a linear cost as in [Davis et al., 1993], and a quadratic cost as in [Moreau et al., 2017]. We chose the linear proportional model over a fixed charge, as we believe that trading costs should scale dynamically according to their size, as in the context of large institutional trading within active OTC derivatives markets, proportional charges are more realistic than fixed charges. We chose not to use the quadratic model due to rebalancing very often, meaning that the changes in position are very small, and thus the quadratic term would be negligible compared to the linear one. The implementation then gives the cost at time t_i as

$$\text{trading cost}(t_i) = c|q_{t_i} - q_{t_{i-1}}|^\top p_{t_{i-1}},$$

for some constant and deterministic trading cost parameter $c \in \mathbb{R}$, where $|v|$ for a vector v means applying the absolute value to each element of v

Thus we want to find some hedging strategy in the set of admissible hedging strategies $\mathscr{Q}$ that solves the optimization problem

$$\underset{q\in\mathscr{Q}}{\text{minimize}}\ \mathbb{E}\left[\sum_{i=1}^{\min(N,\tilde{N})}\mathscr{R}\big((l_{t_i}-l_{t_{i-1}})-(p_{t_i}-p_{t_{i-1}})^\top q_{t_i}\big)-c|q_{t_i}-q_{t_{i-1}}|^\top p_{t_{i-1}}\right]. \quad (3.2)$$

4

Stochastic modeling

This chapter defines the statistical models we will use to build our market simulations. First, we introduce the concepts used in the models, then the specific models themselves. They are all numerically integrated via the Euler–Maruyama method except the Cox–Ingersoll–Ross++ jump model, which is integrated with the balanced implicit scheme. These numerical methods are straightforward to implement and further details about them are omitted as they are not directly relevant to the thesis but are only used to simulate the chosen stochastic models.

4.1 Wiener process

The Wiener process, also known as Brownian motion, is the core of many financial models.

Definition. An $\mathbb{R}$-valued stochastic process $\{W_t\}_{t\geq 0}$ is called a *Wiener process* if the following hold.

i) $W_0 = 0$.

ii) The process has independent increments, i.e. for all $a < b \leq c < d$, the increments $W_a - W_b$ and $W_c - W_d$ are independent.

iii) For each $i < j$, the increment $W_i - W_j$ has a normal distribution with mean 0 and variance $j - i$.

iv) The mapping $t \to W_t$ is continuous for all realizations of the process $\{W_t\}_{t\geq 0}$.

It is used extensively in financial modeling because it is a relatively simple continuous stochastic process with independent increments. This aligns with the standard assumption that past performance gives no information about future movements. Example realizations of a Wiener process are shown in Figure 4.1.

Figure 4.1 Three example realizations of Wiener processes.

4.2 Jump process

For several of our models, we want stochastic discontinuities. This is done by adding a jump process.

Definition. A *jump process* $\{J_t\}_{t\geq 0}$ is a $\mathbb{R}$-valued stochastic Poisson process defined as

$$J_t = \sum_{i=0}^{M_t} Y_i$$

with

i) $J_0 = M_0 = Y_0 = 0$.

ii) M_t a random variable from a time-homogeneous Poisson process with an intensity parameter.

iii) Y_i a random variable from some size distribution.

Example realizations of such a process is shown in Figure 4.2.

4.3 Geometric Brownian motion

A geometric Brownian motion, known for its use in the Black–Scholes model, is an $\mathbb{R}$-valued stochastic process $\{G_t\}_{t\geq 0}$ which satisfies the *stochastic differential equation* (SDE)

$$dG_t = \mu G_t dt + \sigma G_t dW_t,$$

with drift μ and volatility σ. An example of such a process is shown in Figure 4.3.

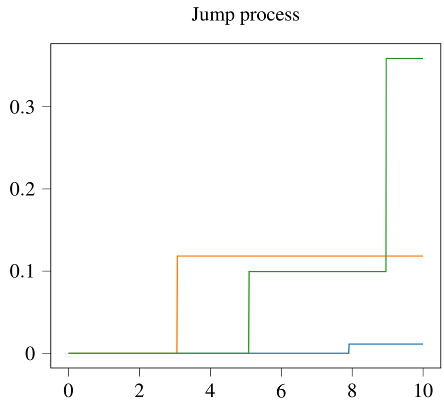

Figure 4.2 Three example realizations of jump processes.

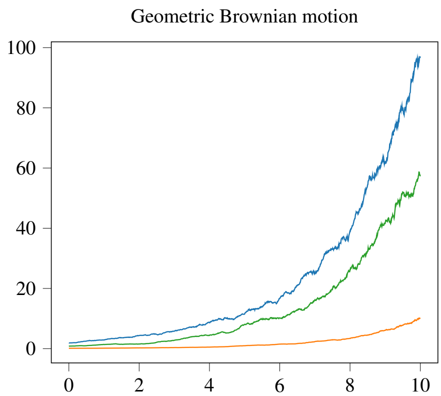

Figure 4.3 Three example realizations of geometric Brownian motions with positive drift.

Geometric Brownian motion without drift

In order to analytically calculate swaption prices we use an unrealistic, though not unused in industry, model for interest rate as a geometric Brownian motion without drift. It is thus represented as simply a process r_t which satisfies

$$dr_t = \sigma r_t dW_t.$$

An example realization of the interest rate as a geometric Brownian motion is shown in Figure 11.7 (sub-figure "interest rate").

Using this model, we introduce pricing formula for a payer swaption (pay fixed rate, receive floating rate). The holder of the swaption has the right to pay a fixed rate K and receive the floating rate r_t on a swap that will last $T_b - T_1$ years starting in $T_1 - t$ years, with payments happening at times T_i for $1 \leq i \leq b$, i.e. annual payments. These rates are derived in the discussion on Swaptions in Chapter 2. Assuming continuous compounding of interest, with some scaling factor based on the notional value L, the pricing formula is:

$$PS(t, T_1, T_b, K, S_{T_1,T_b}(t)) = L\left(\sum_{i=1}^{b} e^{-r(T_i-t)}\right)\left(S_{T_1,T_b}(t)\Phi(d_1) - K\Phi(d_2)\right)$$

where $\Phi : \mathbb{R} \to \mathbb{R}$ is the cumulative distribution function of the standard normal distribution (mean 0 and standard deviation 1),

$$d_1 = \frac{\ln(\frac{S_{T_1,T_b}(t)}{K})\sigma^2(T_1 - t)}{\sigma\sqrt{T_1 - t}}$$

and

$$d_2 = d_1 - \sigma\sqrt{T_1 - t}.$$

Throughout this thesis we assume $L = 1$, as we for simplicity reasons want to study the dynamics of the problem and L is just a scaling factor that does not change results.

4.4 Hull–White model

In the simpler market environment used in the thesis, where we don't need to calculate swaption prices, we instead use the Hull–White model. It is a financial model designed to describe the evolution of interest rates over time. The most important part of it is that it is mean reverting, which is desired for interest rate models. The SDE for the short-term interest rate B_t under the Hull–White model is given by:

$$dB_t = (\tilde{\theta}(t) - aB_t)dt + \sigma dW_t$$

Where:

- B_t represents the short-term interest rate at time t.
- $\tilde{\theta}(t)$ is the time-dependent mean reversion level.
- a is the speed of mean reversion, indicating how quickly the interest rate returns to its mean.
- σ is the volatility of the interest rate process.
- dW_t is a Wiener process representing the random shocks to the interest rate.

The Hull–White model allows interest rates to follow a mean-reverting process, capturing the tendency for interest rates to revert towards a long-term average over time. The Hull–White model is widely used in financial markets for interest rate modeling and derivative pricing due to its flexibility and ability to capture mean reversion dynamics. An example of its behavior is shown in Figure 11.1, where it is designated B_t.

4.5 Merton jump-diffusion model

Merton jump-diffusion models are financial models that extend the geometric Brownian motion model to account for sudden jumps in asset prices. We consider the jump-diffusion model described by the following SDE:

$$dA_t = \mu A_t dt + \sigma A_t dW_t + A_t dJ_t$$

Where:

- A_t represents the probability of default at time t.
- μ is the drift coefficient, representing the expected continuous growth rate of the asset.
- σ is the volatility coefficient, measuring the magnitude of random fluctuations.
- W_t is a standard Wiener process representing continuous random movements.
- dJ_t represents the jump component, where the intensity parameter for the jumps is called λ_J and the size distribution is a normal distribution with mean μ_J and variance σ_J^2.

The jump-diffusion model combines continuous stochastic processes (represented by the drift and diffusion terms) with discontinuous jumps (represented by the jump component). This model was initially created to model stock prices, but it also works well for modeling default probabilities; which is what we will use it for. It builds on the aforementioned geometric Brownian motion, adding jumps to capture the discontinuous movements that can happen when, for example, some shocking news is revealed. An example of its behavior is shown in Figure 11.1, where it is designated A_t.

4.6 Cox–Ingersoll–Ross++ jump-model

Another model we will use for default probability is the Cox–Ingersoll–Ross++ jump-model. Rather than directly giving the probability of default, we will use this to model the default intensity, and then calculate the probability from that, which is a more sophisticated model choice than the Merton jump-diffusion model.

The SDE for the Cox–Ingersoll–Ross++ jump-model is

$$dy_t = \kappa(\mu - y_t)dt + \nu\sqrt{y_t}dW_t + dJ_t.$$

In this model we have for the jump process J_t, that the intensity parameter for M_t is called λ, and the distribution of Y_i is an exponential distribution with mean γ.

To calculate the default probability, we do the following. Begin by denoting the act of default by τ, our probability measure by $\mathbb{Q}$, and an indicator function for some event E by 1_E. We from this model calculate the probability of not defaulting before time T, at time t, as

$$\mathbb{Q}(\tau > T) = 1_{\tau > T}\bar{\alpha}(t,T)\exp(-\bar{\beta}(t,T)y_t).$$

We have here

$$\bar{\alpha}(t,T) = \mathfrak{A}(t,T)\left(\frac{2h\exp\left(\frac{h+\kappa+2\gamma}{2}(T-t)\right)}{2h+(\kappa+h+2\gamma)(\exp(h(T-t))-1)}\right)^{\frac{2\lambda\gamma}{\nu^2-2\kappa\gamma-2\gamma^2}}$$

and

$$\bar{\beta}(t,T) = \mathfrak{B}(t,T);$$

where

$$\mathfrak{A}(t,T) = \left(\frac{2h\exp((\kappa+h)(T-t)/2)}{2h+(\kappa+h)(\exp((T-t)h)-1)}\right)^{2\kappa\mu/\nu^2},$$

$$\mathfrak{B}(t,T) = \frac{2(\exp((T-t)h)-1)}{2h+(\kappa+h)(\exp((T-t)h)-1)},$$

and

$$h = \sqrt{\kappa^2 + 2\nu^2}$$

[Brigo et al., 2013]. An example of how this probability behaves is shown in Figure 11.7 (sub-figure "Probability of Default").

5

Credit risk

This chapter delves into the intricacies of CVA, a pivotal concept in managing the credit risk inherent in OTC derivatives. We begin by exploring some essential concepts associated with CVA, then introduce the motivation behind CVA and its fundamental definition. The discussion then advances to the mathematical models we use for modeling CVA. We introduce two different models. The first is a simple one but still captures significant characteristics. The second model is the pricing model proposed by [Brigo et al., 2013]. It is more complex than the first one and is also more representative of real CVA derivatives in financial markets.

Cross-gamma

Cross-gamma appears when the process one wants to hedge is non-linear in its underlying assets. It is a second-order derivative that measures the sensitivity of a derivative's price, not to the change of one underlying asset, but to the simultaneous changes in two different underlying assets. Technically, it is defined as: given a derivative ϖ defined on some underlying assets $S_1, \ldots S_n$, the cross-gamma is

$$\left\{ \frac{\partial^2 \varpi}{\partial S_i \partial S_j} \right\}_{i \neq j}.$$

This interaction is critical in the context of hedging because it affects the efficacy of simple delta hedging strategies. When there is no cross-gamma and no market frictions, hedging is quite easy, but when it is there, even without other frictions, it becomes very difficult. Delta-hedging alone (which only considers the first-order derivatives with respect to changes in underlying assets) might not be sufficient and could lead to suboptimal hedging outcomes when not rebalancing continously, potentially resulting in losses due to not accounting for interactions between multiple assets.

Wrong-way risk

Wrong-way risk (WWR) is the additional risk associated when the portfolio being hedged and the probability of counterparty default "are 'correlated' in the worst possible way" [Brigo et al., 2013]. For example, if the portfolio is a big loan with a floating interest rate, and the counterpart is doing badly, then interest going up makes the probability of them defaulting due to not being able to pay go up, while the value of the contract to you also goes up as they should pay more. In other words, the potential losses given that they default go up as their probability of defaulting rises. There is another closely related term used when the risk goes in the "opposite" direction, namely *right-way risk* (RWR). This happens when potential losses are negatively correlated with default probability. It is worth mentioning that the correlation of such processes is not necessarily constant over time, however, the simulations conducted in this thesis are limited to constant correlation. In industry practices, these risks are categorized as general and specific risks, where general considers general market risk factors and specific considers individual counterparty factors. For simplicity, only the general category is considered.

5.1 Credit valuation adjustment

The CVA for a derivative is calculated as the discrepancy between its risk-free value, assuming there is no counterparty credit risk, and its actual market value:

$$\mathrm{CVA} \triangleq PV_{\text{risk-free}} - PV,$$

where PV is the present value of the considered derivative contract, making the CVA a market price for the counterparty credit risk [Zhu and Pykhtin, 2007]. The models used to price these adjustments depend on the considered derivative contract, so different types of swaps have slightly different pricing models that account for the specific underlying factors in these trades. The idea is still the same for all though, namely to fairly account for the counterparty credit risk [Brigo et al., 2013].

We use the risk-neutral pricing methodology as described in [Brigo et al., 2013]. We will only be interested in a unilateral CVA, which means that the investor (typically a sell-side Bank) is assumed to be default-free, only the counterparty in the contract can default. We consider the basic case where the CVA trade is not supported by any collateral. We also assume that we have only one contract, so our portfolio is only the contract and our hedge, and there is only one counterparty.

Regulation challenges

Regulators seek to standardize the methodology for assessing CVA risk, but its complexity and reliance on detailed modeling pose challenges for accurate valuation. The need for multiple simulation levels limits the scenarios that can be feasibly analyzed, and while some methods exist to bypass these limitations, they often require assumptions that may compromise the accuracy of CVA estimates, particularly affecting the assessment of wrong-way risk [Brigo et al., 2013].

Credit default swap

In many ways, the most important tool for dealing with CVA is the *credit default swap* (CDS). The most basic form of a CDS is quite simple, it acts like an insurance. There is a buyer and a seller, and the buyer periodically pays the seller for a predetermined number of years, in exchange for a big payment from the seller if some reference entity defaults [Bomfim, 2022]. CDS:s are by far the most common type of credit derivative [Bomfim, 2022]. Specific CDS contracts are still, however, quite illiquid [Bomfim, 2022]. This poses a large problem for using them as hedging devices.

They are used in the context of CVA as a way to, by proxy, measure the default risk. Practically, this is done by calibrating the parameters in the default intensity model to the CDS prices [Brigo et al., 2013]. Since they are such a good proxy for the default risk, they are also the primary instrument used to hedge that portion of a CVA. As a result of this close relationship, we will not be simulating both the default probability and CDS:s, instead we will assume that the agent can trade directly in the default probabilities when building its hedge.

5.2 Simple CVA model (market environment 1)

Our initial choice of a model for the CVA is a price process of the form $CVA_t = A(t) \cdot B(t)$. We choose this as the first model to make as it captures most complexities of CVA, whilst only containing two underlying assets.

The way to interpret this as a CVA is to think of one of the processes as the default probability and the other as the expected future losses given default. One can think of it as a CVA with on a contract with only one payment, very far into the future. This is quite a big simplification, as it reduces an integral into a single product, however, it is still surprisingly useful. Most importantly, it captures the cross-gamma and wrong-way risk effect.The market we will simulate will thus consist of the two underlying stochastic processes $A(t)$ and $B(t)$, which can be correlated. An example of how this market behaves is shown in Figure 11.1.

$A(t)$ will follow a Merton jump diffusion model, and $B(t)$ will follow a Hull–White model, models described in Chapter 4. Oftentimes the market is said to also include a bank, as a risk-free investment. For many derivatives, this is necessary to hedge fully. For a bank, however, it is oftentimes strongly preferable to not have money just sitting, and so when possible they wish to hedge without such a process.

The Merton jump-diffusion model was chosen as default probabilities often have jumps, so it is important to include those, and by choosing parameters such that the net drift of the model is negative we can know that as time passes the default probability goes to zero, which we want as the probability should be zero at the end of the contract. The Hull–White model was chosen as, if there is only one payment in the contract and it is placed at the end, then we expect that the expected losses given default should remain around the same mean the entire time, and oftentimes these will depend on interest rate so we should expect this process to behave similarly to the interest rate; thus it is quite fitting to use a mean-reverting model commonly used to model interest rates. The specific parameters used for the models are presented in Table 5.1. We arrived at these values by comparison with real markets and with guidance from NORDEA, however, they can be made more realistic by properly calibrating them to market data.

Table 5.1 Market environment 1 parameters.

Parameter	**Value**
Shared parameters	
Time step (Δt)	$1/252$
Hull–White model	
Initial rate (r_H)	0.3
Volatility (σ_H)	0.03
Jump-diffusion model	
Initial probability (S_0)	0.4
Volatility (σ_D)	0.05
Drift (μ)	0.01
Jump intensity (λ_J)	0.1
Jump mean (μ_J)	0.1
Jump standard deviation (σ_J)	0.5

5.3 IRS CVA model (market environment 2)

Here we will present the second CVA model we consider, the fair price model for a CVA of an *interest rate swap* (IRS). We consider the payer IRS where we pay a fixed interest rate and receive a floating interest rate. This model is chosen as it is much more realistic, and thus also much more complex, than the previous one.

Assume the swap begins at time T_a and ends at time T_b. Then the CVA can be priced as

$$\begin{aligned}\mathrm{CVA}_t &= \text{L}_{\text{GD}}\mathbb{E}_t\left[\mathbf{1}_{\{\tau \le T_b\}} D(t,\tau)\max(\mathrm{PV}(\tau),0)\right] \\ &= \text{L}_{\text{GD}}\int_{T_a}^{T_b} PS(t,s,T_b,K,S_{s,T_b}(t))d_s\mathbb{Q}\{\tau \le s\}\end{aligned} \tag{5.1}$$

[Brigo et al., 2013]

where:

- CVA_t is the credit valuation adjustment at time t.
- LGD is the loss given default, representing the percentage of the exposure that is lost if a default occurs.
- $\mathbb{E}_t$ denotes the conditional expectation given the information available at time t.
- $\mathbf{1}_{\{\tau \le T_b\}}$ is an indicator function that is 1 if the default time τ occurs on or before the maturity T_b, and 0 otherwise.
- $D(t,\tau)$ is the discount factor from time t to the default time τ.
- $\mathrm{PV}(\tau)$ is the present value of the contract at the default time τ.
- T_a and T_b are the start and end times of the contract.
- $PS(t,s,T_b,K,S_{s,T_b}(t))$ is the price of a swaption that allows entering into a swap at time s that matures at T_b, with a strike rate K and a swap rate $S_{s,T_b}(t)$.
- $\mathbb{Q}\{\tau \le s\}$ is the probability of default by time s.
- $d_s\mathbb{Q}\{\tau \le s\}$ indicates integrating over the probability measure $\mathbb{Q}\{\tau \le s\}$ as s varies over the time interval from T_a to T_b.

In practice the loss given default, LGD, is often set as a constant 0.6, meaning you expect to lose 60% of the remaining value if they default. We omit the LGD going forward as it depends on the specific contract/counterparty and may require knowledge about that specific trade, and is often constant over time.

Since we only get money at discrete time points, we can break this integral up into a sum. Assuming then that a default in the interval $(T_i, T_{i+1}]$ means losing out on all payments after and including T_{i+1}, the integral (5.1) becomes:

$$\begin{aligned}\mathrm{CVA}_t &= \sum_{i=a+1}^{b} PS(t, T_i, T_b, K, S_{T_i,T_b}(t))\mathbb{Q}\{\tau \in (T_{i-1}, T_i]\} \\ &= \sum_{i=a+1}^{b} PS(t, T_i, T_b, K, S_{T_i,T_b}(t))(\mathbb{Q}\{\tau > T_{i-1}\} - \mathbb{Q}\{\tau > T_i\})\end{aligned}$$

[Brigo et al., 2013].

We specifically model the CVA process for an IRS over 10 years with yearly payment. We will use the geometric Brownian motion without drift as a model for the interest rate, and the Cox–Ingersoll–Ross++ jump-model for the default probability. For the Cox–Ingersoll–Ross++ jump-model we have 6 parameters we must give values. Except for λ and γ, they are taken from [Brigo et al., 2013], for good reasons. Those we had to tune ourselves to get reasonable behavior. The values for the geometric Brownian motion without drift were chosen entirely by us. The model parameters are shown in Table 5.2. An example of how this market behaves is shown in Figure 11.7.

Table 5.2 Market environment 2 parameters.

Parameter	**Value**
Shared parameters	
Time step (Δt)	$1/252$
Time interval ($\tilde{\beta}$)	1
Initial time (T_A)	0
Final time (T_B)	10
Geometric Brownian motion without drift	
Initial rate (r)	0.03
Volatility (σ)	0.1
Cox–Ingersoll–Ross++ jump-model	
Initial intensity (y_0)	0.035
Volatility (ν)	0.15
Speed of adjustment to mean(κ)	0.35
Mean (μ)	0.045
Jump related intensity (λ)	0.001
Jump related mean (γ)	0.0005

6

Hedging CVA

This chapter explores the various metrics used to evaluate *profits and losses* (P&L), emphasizing the importance of measuring P&L variance and other risk metrics. We will define the key variables involved in calculating P&L for a hedged portfolio and introduce benchmarks and performance metrics that will be used to compare different hedging strategies.

Additionally, we will delve into our environments to illustrate how these concepts apply there, explaining how to manage and mitigate financial risks through effective hedging by allocating in the underlying instruments.

6.1 Profits and losses

In financial mathematics research, P&L is one of the main factors evaluated in hedging and investment strategies [Esipov and Vaysburd, 1999]. There are several metrics that P&L can be evaluated on, we will discuss these metrics in this section, highlighting the importance of measuring the variance of P&L [Berns, 2014] and other risk metrics.

P&L is the change in the value of a portfolio over time. It is the main metric for evaluating a hedge. A theoretically optimal hedge has profits and losses constantly equal to zero, although in practice the goal is often not to reduce risks to zero but to lower them to an acceptable margin and in such cases the the P&L will vary over time.

Let's define the variables involved in calculating the P&L of a arbitrary portfolio with a hedge[1]:

- $P_{\text{contract}}(t)$: Price of the contract at the time of sale.

[1] The P's and Q's can also be seen as vectors, with each entry representing one contract/hedging instrument

- $P_{\text{hedge}}(t)$: Price of the hedge at the time of sale.
- $Q_{\text{contract}}(t)$: Quantity of the contract.
- $Q_{\text{hedge}}(t)$: Quantity of the hedge.
- $P_{\text{contract}}(T)$: Price of the contract at a later time (e.g., at P&L settlement[2]).
- $P_{\text{hedge}}(T)$: Price of the hedge at P&L settlement.

The P&L under some time becomes the following:

$$\text{P\&L}(T,t) = (P_{\text{contract}}(T) - P_{\text{contract}}(t)) \cdot Q_{\text{contract}}(t) - (P_{\text{hedge}}(T) - P_{\text{hedge}}(t)) \cdot Q_{\text{hedge}}(t) \tag{6.1}$$

You can recognize this as exactly the input to our risk measure in (3.1), assuming no trading cost.

6.2 Benchmarks

To evaluate our models, we will be comparing them against two other strategies. These will be delta hedging, and doing nothing.

Delta hedging

The "delta", in the context of financial derivatives, is one of the sensitivities of that derivative. It indicates how the price of a derivative changes in response to an underlying. Specifically, delta measures the rate at which the price of a derivative changes for every one unit change in the price of the underlying asset.

Assume we have an arbitrary derivative $\varpi(S_1(t),..,S_n(t),t)$, whose value depends on some underlying assets $\{S_i(t)\}_{i=1}^n$. Then the deltas with respect to each underlying asset S_i to are $\frac{\partial\varpi}{\partial S_i}$. We then create a hedge out of these deltas, which we call the delta hedge $\delta_\varpi(t)$, by

$$\delta_\varpi(t) = \sum_{i=1}^n \frac{\partial\varpi}{\partial S_i} S_i(t).$$

This can be seen as taking the first order Taylor expansion of the contract, thus guaranteeing similar behavior locally. Delta hedging is used to offset the delta of a position in a derivative by taking an opposite position in the underlying asset. The

[2] P&L reporting frequency can vary by institution, regulatory requirements, and the specific needs of the business. In our case, we are interested in daily P&L i.e. ($T = t + \tilde{\Delta t}$), where t has time unit years, and $\tilde{\Delta t}$ is one day.

goal is to achieve a delta-neutral position, meaning the overall delta of the portfolio is zero. This results in optimal hedging, assuming that the process being hedged is continuous and that the hedge is being continuously rebalanced. Unfortunately, it does not take into account trading costs at all, and so it becomes very expensive when they exist. We use delta hedging as a benchmark because it is very simple to compute given an equation for the price of the derivative, and it performs very well. As a result of using constant volatility in our underlying models, the minimum variance hedge defined in [Hull and White, 2017] is equivalent to delta hedging. We expect this to perform extremely well with no trading cost, to be slightly worse but still good with low trading cost, and to be quite bad at high trading cost.

Doing nothing

"Nothing" means exactly to not attempt to hedge. We should always expect that when the trading costs are low (0 or 5 percent) our models should outperform "nothing". On the other hand, at an unreasonably high trading cost of 100 percent, we expect "nothing" to do very well. We also expect "nothing" to do comparably well when the processes are strongly negatively correlated, as in those cases they, to some extent, hedge themselves.

6.3 Performance metrics

As we rebalance at a rate of once a day, we get a $P\&L$ value for every day. We denote this by $P\&L(t, t-\Delta t) = P\&L_t$. Taking the negative of the $P\&L$ to convert it to losses, we get a series $L_t = -P\&L_t$. The distribution ℓ is taken from the empirical distribution of the time series L. We do this conversion so that a lower value is better for all our metrics; as a better hedge should lower risk, thereby lowering losses.

Variance of losses

This is simply the empirical variance of the losses of our strategy. It is very important as even if a strategy had mean 0 losses, if the variance is high there can be a large risk of large losses. Thus one would prefer a strategy with non-zero, though very small, mean losses but with low variance over one with zero mean but high variance.

Variance of losses excluding first point

This one is important because when there is trading cost, the initial cost of taking a hedging position massively increases the empirical variance. Thus this is a more fair valuation for the variance of the losses throughout the use of the strategy.

Mean square loss

Least squares is the most common metric used for minimizing. In this case it practically means penalizing losses and profits equally.

Mean loss

As profits are not necessarily bad, it can be interesting to look at metrics which reward them. This is one such metric. In this case a negative mean loss means a mean profit, which is a good thing.

Value at risk

Value at risk (VaR): Given a confidence level α, the Value at risk at α for a loss distribution ℓ over a time period T is defined as:

$$\mathrm{VaR}_\alpha(T) = -\inf\{l \in \mathbb{R} : F_\ell(l) \geq \alpha\}$$

where F_ℓ is the cumulative distribution function of losses. See Figure 6.1 for a simple illustration of this metric.

This metric answers the question: What is the minimum loss over the whole range of outcomes in the considered tail?

Similarly to the variance, minimizing this is more about minimizing the risk of large losses, as opposed to average losses.

Expected shortfall

Expected shortfall (ES): For the same setup as the **Value at risk (VaR)**, the Expected shortfall at confidence level α is the conditional expectation of losses exceeding the VaR:

$$\mathrm{ES}_\alpha(T) = \mathbb{E}[l | l \leq \mathrm{VaR}_\alpha(T)]$$

This metric answers the question: What is the average loss over the whole range of outcomes in the considered tail?

Minimizing this one again minimizes the risk of large losses.

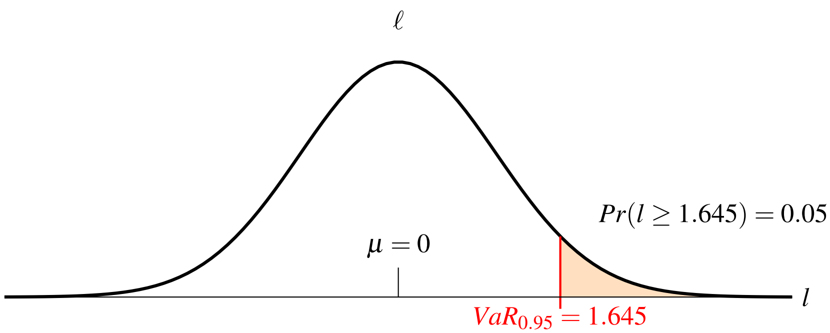

Figure 6.1 Illustration of VaR_α, for $l \sim \ell = N(0,1)$ with $\alpha = 0.95$.

Turnover

This is a measure of how much trading is being done. There is no law for if it is good or bad to have low or high, it depends on what your purpose is. In our scenarios with trading costs though, it is best kept low, as trading costs make it more expensive to trade. It is defined as

$$Turnover = 100\frac{\min(total\ purchases, total\ sales)}{Average\ portfolio\ value}$$

Note! We will refer to "Variance of losses" and "Variance of losses excluding the first point" as the variance metrics, and we will call the "mean squared losses", the "Value at risk", and "expected shortfall" the risk metrics for convenience.

6.4 P&L in environment 1

For our process $P_t = A_t \cdot B_t$, we consider the hedge $H_t = \beta_t A_t + \alpha_t B_t$, where we are allocating β_t respective α_t in quantity of our underlying processes A_t respective B_t at time $t - \Delta t$. These quantities should be adjusted as to be able to minimize the risk through time. For that, it is required to analyze the P&L.

The P&L becomes (as in 6.1):

$$\begin{aligned} \text{P\&L}_t &= \Delta P_{\text{contract}}(t) - (\Delta P_{\text{hedge}}(t) \cdot Q_{\text{hedge}}(t)) \\ &\overset{\text{Environment 1}}{=} (P_t - P_{t-\Delta t}) - (\beta_t \cdot (A_t - A_{t-\Delta t}) + \alpha_t \cdot (B_t - B_{t-\Delta t})) \\ &\overset{\text{simplified notation}}{=} (\Delta P) - (\beta_t \cdot \Delta A + \alpha_t \cdot \Delta B) \end{aligned} \tag{6.2}$$

where $\Delta t = \frac{1}{252}$ as t is in units years, and we want daily P&L.

6.5 P&L in environment 2

The P&L is still defined as in the general case 6.1. However, now we are in another environment setting. We are no longer hedging in only 2 processes/contracts (i.e. in Environment 1, the A and B processes). We instead have more standard underlying products to hedge in, i.e. the underlying default probabilities[3] and Swaptions. Considering a maturity of 10 years for the underlying interest rate swap, with yearly payments, we have then $2 \cdot 10 = 20$ underlying products to hedge under the tenor of the CVA (see 5.3), i.e.

$$H_t = \sum_{i=a+1}^{b} (\beta_t^{T_i} \cdot \mathbb{Q}\{\tau \in (T_{i-1}, T_i]\}) + (\alpha_t^{T_i} \cdot PS(t, T_i, T_b, K, S_{T_i,T_b}(t))),$$

[3] see discussion on CDS in Section 5.1

where $a = 0$ and $b = 10$, and $\beta_t^{T_i}$ and $\alpha_t^{T_i}$ mean the quantity held from time $t-\Delta t$ to time t for the underlying assets corresponding to time interval $(T_{i-1}, T_i]$. So the P&L becomes (as in 6.1 and 6.2):

$$
\begin{aligned}
\text{P\&L}_t &= \Delta CVA - (\Delta \mathbf{H} \cdot \mathbf{Q}_{\text{hedge}}(t)) && (6.3)\\
\mathbf{Q}_{\text{hedge}_t} &= \begin{bmatrix} \beta_t^{T_1} & \beta_t^{T_2} & \cdots & \beta_t^{T_{10}} & \alpha_t^{T_1} & \alpha_t^{T_2} & \cdots & \alpha_t^{T_{10}} \end{bmatrix}^T && (6.4)\\
\mathbf{v} &= \Delta \begin{bmatrix} \mathbb{Q}\{\tau \in (T_0, T_1]\} & \mathbb{Q}\{\tau \in (T_1, T_2]\} & \cdots & \mathbb{Q}\{\tau \in (T_9, T_{10}]\} \end{bmatrix}\\
\mathbf{u} &= \Delta \begin{bmatrix} PS(t, T_1, T_{10}, K, S_{t,T_{10}}(t)) & \cdots & PS(t, T_{10}, T_{10}, K, S_{T_{10},T_{10}}(t)) \end{bmatrix}\\
\Delta \mathbf{H} &= \begin{bmatrix} v & u \end{bmatrix}
\end{aligned}
$$

6.6 Remark on P&L, hedging, and frictions

Mathematically, the goal of a hedge is to minimize the downside of P&L. Returning to the model from environment 1, and assuming we buy a position of β units of A and α units of B at time $t-dt$ in order to hedge P, we have

$$\text{P\&L}_t = (P_t - P_{t-dt}) - (\beta_t(A_t - A_{t-dt}) + \alpha_t(B_t - B_{t-dt})).$$

Here the time increments dt represent how often we can rebalance the hedge. We see now that by dividing by dt and allowing dt to go to zero we get $\frac{\text{P\&L}}{dt} = \frac{dP}{dt} - (\beta_t \frac{dA}{dt} + \alpha_t \frac{dB}{dt})$, but we know from the definition $P_t = A_t \cdot B_t$ that $\frac{dP_t}{dt} = B_t \frac{dA}{dt} + A_t \frac{dB}{dt}$. Thus this is zero only when $\beta_t = B_t$ and $\alpha_t = A_t$, and we can recognize this as exactly delta hedging. In general, a similar argument shows that delta hedging is always optimal in continuous settings.

If we add frictions, however, the ideal hedge is no longer so obvious. Adding just linear trading costs, with parameter c, assuming we before buying our new position held β_{t-dt} units of A and α_{t-dt} units of B but once again at time $t-dt$ bought enough to have β_t units of A and α_t units of B, we get $\text{P\&L}_t = (P_t - P_{t-dt}) - (\beta_t(A_t - A_{t-dt}) + \alpha_t(B_t - B_{t-dt}) - c(|\beta_t - \beta_{t-dt}|A_t + |\alpha_t - \alpha_{t-dt}|B_t)$ once again dividing by dt and letting it go to zero we get $\text{P\&L}/dt = \frac{dP}{dt} - (\beta_t \frac{dA}{dt} + \alpha t \frac{dB}{dt}) - c(\frac{d|\beta_t|}{dt} A_t + \frac{d|\alpha_t|}{dt} B_t)$. Clearly now delta hedging is not going to be the ideal solution, however finding the ideal solution analytically is intractable.

7

Reinforcement learning

This chapter will define everything we need related to reinforcement learning. First, we explain the setup necessary for reinforcement learning to be applied to a problem, and then we define the algorithms used to solve the problem. Finally, we present the reward functions that our reinforcement learning algorithm will learn from.

7.1 Reinforcement learning

Reinforcement learning (RL) is a subset of machine learning that does not use a data set as a starting point, instead generating data based on the needs of the optimization algorithm [Bertsekas, 2023]. It is especially applicable to our problem due to the problem's sequential nature.

Markov decision process

The ideal situation to apply reinforcement learning to is a Markov decision process. This is a process within which the next state depends solely on the current state, and possibly the current action.

Our situation is exactly a Markov decision process. In fact, the current action doesn't even affect the next state. This follows from the definitions of our markets. The defining SDEs depend only on terms from time t, and the random increments are independent. This allows us to very naturally define the possible actions we can make and states we can observe.

Action space

The action space is the space from which the agent can take actions [Bertsekas, 2023]. In order to build a proper hedge, our agent should be able to trade arbitrary amounts of each underlying asset. Thus the action space should be a space allowing it to assign a real number for each asset.

Observation space

The observation space is the space that the observations the agent gets come from [Bertsekas, 2023]. As we want our agent to be able to trade in all underlying assets and take into account trading costs, it needs to see the values of the contract it is trying to hedge, the values of the underlying assets, and how much of each of those assets it is currently holding.

Reward

To learn from Markov decision processes, one must have some well-defined notion of which states one wants to be in or what it means for an action to be good in a given state. To do this, a reward function is defined. This function can vary significantly in how it looks. If, for example, one is trying to train an agent to play chess, one might give no reward until the end of the game, then a positive reward if the agent won and a negative reward if they lost. This is very risky, though, as there are so many actions taken between any given action and the reward, and so it is difficult for the agent to learn which of the moves it made was the bad one if it loses. Thus, if possible, one wants to give rewards more continuously. In our case, as we want to evaluate our models based on P&L, and we get a daily value of P&L, by constructing reward functions based on the P&L, we can give rewards to our agents every time they make an action. The way the interaction between the environment giving updated states and rewards to the agent and the agent giving actions to the environment looks is shown in Figure 7.1.

When evaluating which move to make, one wants to look at the expected rewards in the future. Because one is less sure about the state and reward reached the further into one looks, one often has a discount factor that makes more distant expected rewards worth less.

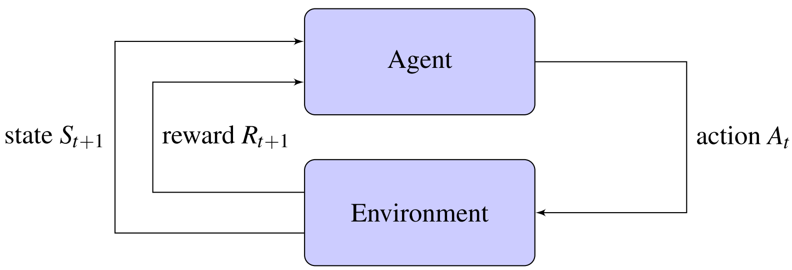

Figure 7.1 Diagram of interaction between agent and environment in reinforcement learning

Policy, and value functions

The action to be taken by an agent is defined by the policy function. It is a function taking in a state and possible action, and returning the probability with which one should select that action. It is generally denoted as $\pi(a|s)$ where a is the action whose probability one wants to find out and s is the current state. The goal of training an agent is to find the correct policy function.

In order to find the policy function, it can be useful to know which states are good. How good a state is, given that you will act according to a given policy, is given by the value function. With R_{t+k+1} as the reward given at time $t+k+1$ for acting according to the policy, ω as a discount factor and $S_t = s$ as the current state, the value function is defined by

$$v_\pi = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \omega^k R_{t+k+1} \middle| S_t = s \right].$$

Policy gradient methods

Policy gradient methods are a subset of RL methods which optimize directly for the policy function, as opposed to optimizing it by proxy through finding the correct value function. This is done through parametrizing the policy function, then performing gradient ascent [Sutton and Barto, 2018]. That is, with A_t as the action taken at time t, S_t as the state at time t, and policy parameters $\hat{\theta}_t$ at time t, defining our policy function as

$$P(A_t = a | S_t = s, \hat{\theta}_t = \hat{\theta}) = \pi(a|s, \hat{\theta})$$

the rule for updating the parameters is

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \hat{\alpha} \widehat{\nabla J}(\hat{\theta}_t),$$

where $\widehat{\nabla J}(\hat{\theta}_t)$ is a stochastic approximation of the gradient of a given performance metric J, and $\hat{\alpha}$ is a variable controlling how large updates can be made.

Actor-critic methods

The previous section has an issue, which is that the performance metric J is unknown. Actor critic methods solve this by letting J depend on the value function, and learning both the policy function and the value function concurrently [Sutton and Barto, 2018].

The optimization is performed through iteration of the two steps:

- Simulate actions by current actor (policy function), and update the critic(value function)

- Improve current Actor by gradient ascent using the newly updated critic

[Bertsekas, 2019]. In deep reinforcement learning, which is how we implement it, they are both represented by different neural networks.

Proximal policy optimization

For the reinforcement learning model, we will use the actor-critic model *proximal policy optimization* (PPO); as [Du et al., 2020] found that it is at least as good and faster to train compared to other deep reinforcement learning models such as DQL. It has also been shown to outperform other policy gradient methods in various standard benchmark tasks, including simulated robotic locomotion and Atari game playing. It strikes a favorable balance between sample complexity, simplicity, and wall-time, making it suitable for this task [Schulman et al., 2017].

Instead of value function we look at what is called the advantage function. It measures how much better some given set of actions is relative to others on average. With a trajectory of length T, a discount factor of ω, and some constant ξ, it is defined as

$$\hat{A}_t = \delta_t + (\omega\xi)\delta_{t+1} + \cdots + (\omega\xi)^{T-t+1}\delta_{T-1} \tag{7.1}$$

$$\text{where } \delta_t = R_t + \omega v(s_{t+1}) - v(s_t), \tag{7.2}$$

with R_t the reward at time t and v the value function [Schulman et al., 2017]. The performance metric we try to optimize for is defined as

$$J_t(\hat{\theta}) = \hat{\mathbb{E}}_t\left[L_t^{CLIP}(\hat{\theta}) - c_1 L_t^{VF}(\hat{\theta}) + c_2 S[\pi_{\hat{\theta}}](s_t)\right] \tag{7.3}$$

where c_1 and c_2 are constants, S is an entropy bonus,

$$L_t^{VF}(\hat{\theta}) = (v_{\hat{\theta}}(s_t) - v_t^{targ})^2 \tag{7.4}$$

is a squared error loss function of the value of the policy against a target value v_t^{targ}, and

$$L_t^{CLIP}(\hat{\theta}) = \hat{\mathbb{E}}_t\left[\min(R_t(\hat{\theta})\hat{A}_t, \text{clip}(R_t(\hat{\theta}), 1-\varepsilon, 1+\varepsilon)\hat{A}_t\right] \tag{7.5}$$

where the clip function clips the probability ratio, so that there is no value in moving outside the interval $[1-\varepsilon, 1+\varepsilon]$ [Schulman et al., 2017]. To train one iterates running the policy for some number of steps, computes the advantage estimates for each of those steps, then for each epoch one divides up the actions from the policy into minibatches and optimizes the performance metric by updating the parameters. This is shown in pseudo-code in Algorithm 1.

As we are focused on the application of RL to CVA models and not the RL algorithm on their own, we use an open source implementation of PPO from Stable

---

Algorithm 1 PPO Training Algorithm [Schulman et al., 2017]

Require: Number of iterations I, steps T, minibatch size M, epochs K and clipping size ε

for iteration in $1 \dots I$ **do**
    Run policy $\pi_{\hat{\theta}_{\text{old}}}$ in environment for T steps
    Compute advantage estimates $\hat{A}_1 \dots \hat{A}_T$
    for epoch in $1 \dots K$ **do**
        Optimize J w.r.t. $\hat{\theta}$ with minibatch size $M \leq T$
    end for
    $\hat{\theta}_{\text{old}} = \hat{\theta}$
end for

Baselines3, which is essentially an improved version of the algorithm presented in [Schulman et al., 2017] which was developed in OpenAI. The relationship is shown in Figure 7.2.

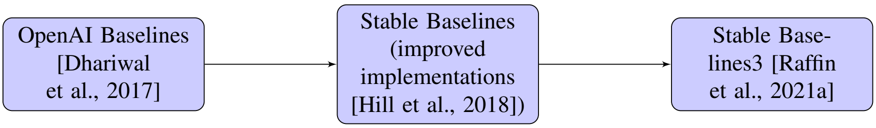

Figure 7.2 Stable Baselines3 is an updated version of Stable Baselines, which is a set of improved implementations of OpenAI Baselines

Exploration and Exploitation

A pivotal concept of learning in RL is the exploration/exploitation trade-off. This trade-off involves managing the balance between exploring the environment and exploiting existing knowledge about the environment. Exploration involves investigating the environment through random actions to gather more information about it. Exploitation, on the other hand, focuses on using known information to maximize rewards [Hugging Face, n.d.]

Different algorithms have different guidelines that dictate the balance between exploration and exploitation to effectively navigate this trade-off. The success of these algorithms depends on the environment and problem they are operating on. The PPO algorithm employs an on-policy method to develop a stochastic policy, actively engaging in exploration by executing actions determined by the latest update of its stochastic policy. The degree of randomness in these actions is influenced by the starting conditions and the training process. As training progresses, the policy generally becomes more deterministic, driven by an update rule that favors exploiting previously discovered rewards; however, this tendency can lead the policy to become stuck in a local optimum [Raffin et al., 2021b].

7.2 Model setup

Environment 1 action space

We have only the two underlying processes, $A(t)$ and $B(t)$. Thus our action space will be

$$\mathscr{A}_{\text{Env }1} = \mathbb{R}^2.$$

Environment 1 observation space

In this case, we have only the two underlying processes $A(t)$ and $B(t)$. As A represents a probability, it must be within the interval $[0,1]$. The interest rate, CVA, and how much underlying we can buy are all real numbers. Thus our observation space will be

$$\mathscr{O}_{\text{Env }1} = \mathbb{R}^4 \times [0,1].$$

Environment 2 action space

Given a swap with n payments, we have $2n$ underlying assets: n swaptions, and n default probabilities. Thus our action space will be

$$\mathscr{A}_{\text{Env }2} = \mathbb{R}^{2n}.$$

Environment 2 observation space

In this case, given a swap with n payments, we have $2n$ underlying assets, n swaptions and n default probabilities. Once again, all default probabilities lie in $[0,1]$. Also, once again, how much underlying we can buy are all real numbers. However, the interest rates and CVA are limited to positive real numbers due to geometric Brownian noise being positive if the initial value is positive. Thus our observation space will be

$$\mathscr{O}_{\text{Env }2} = \mathbb{R}^{2n} \times \mathbb{R}_+^{n+1} \times (0,1)^n.$$

Reward functions

These reward functions are defined to measure the effectiveness of trading decisions based on P&L calculated for a given state and action. The "Current State" refers to the current market conditions and portfolio state, while "Current Action" refers to the trading action being evaluated. The first five metrics use daily observed P&L, and the last two metrics use a function of a 30-days window of P&L. Definitions for the functions **VaR** (value at risk) and **ES** (expected shortfall) in section 6.3.

Reward Function 1.

$$R_1(s_t, a_t) = -|\text{P\&L}|$$

This reward function aims to keep P&L as close to 0 as possible. This is useful when aiming for as neutral portfolio as possible, i.e. losses should be avoided as much as possible and profits are not considered the main goal. In other words, this reward function penalizes both profits and losses the same amount, which might seem controversial, but it might still fulfill the main objective i.e. neutral position.

Reward Function 2. This function rewards profits as much as it penalizes losses. Thus it, relative to the other rewards here, strongly incentivizes profit-seeking.

$$R_2(s_t, a_t) = \text{P\&L}$$

Reward Function 3. This function penalizes losses but ignores profits. Thus it is encouraging behavior somewhere in-between the previous two rewards, not actively trying to avoid profiting, but not trying to prifit if it would mean increased risk of losses.

$$R_3(s_t, a_t) = \min(\text{P\&L}, 0)$$

Reward Function 4. Similarly to Reward function 1, this function penalizes all non-zero P&L, aiming for the most "neutral" hedging strategy as possible. The main difference is that this is a differentiable function, whereas reward function 1 is not.

$$R_4(s_t, a_t) = -(\text{P\&L})^2$$

Reward Function 5. For this metric, we simulate the price process for $N = 10^4$ paths one time-step ahead, and look at the variance of the P&L obtained for these paths. The agent would act in a way that minimizes the variance, regardless on how the price process evolves. This is very different to the other reward functions due to it's dependence on access to the market model, using it to simulate the paths.

$$R_5(s_t, a_t) = -\text{Var}\left(P\&L\right))$$

Reward Function 6. This reward function, value at risk, is a standard risk metric implemented in risk management and risk regulations. What we see below is that our agent doesn't get any reward until we obtain 30 days time series of the P&L. Then we obtain a 30-days window iteratively of P&L, that we use to estimate the **VaR**. It gives a measure of the largest losses we could expect at a given probability, given normal market conditions.

$$R_6(s_t, a_t) = \begin{cases} 0 & \text{if } t < 30 \\ \text{VaR}_{99\%}(\text{monthly P\&L}) & \text{otherwise} \end{cases}$$

Reward Function 7. Similarly to reward 6, this function, expected shortfall, is also a standard tool in risk management and risk regulations. It is also implemented very similarly, taking in a 30 day window of P&L values, and returning a measure of the expected returns in the worst cases at a given percent likelihood. The advantage to value at risk is that it is more sensitive to the tails of the P&L distribution.

$$R_7(s_t, a_t) = \begin{cases} 0 & \text{if } t < 30 \\ \text{ES}_{99\%}(\text{monthly P\&L}) & \text{otherwise} \end{cases}$$

8

Numerical simulations and results

In this chapter we begin by stating the exact setup we use for training our reinforcement learning models. We then present the results of simulations comparing our trained models to the benchmarks.

8.1 Market parameters

All simulations are done for 10 years or until the probability of default hits 1, with 1 year containing 252 days. The market parameters presented in sections 5.2 and 5.3 for environments 1 and 2 respectively are held constant, though we add that we vary two parameters: the correlation ρ between the wiener processes in our market models and the trading cost. This allows us to see if the RL model can adapt its hedging strategies to trading costs and wrong- and right-way risk.

Correlation

As stated in Section 5, financial processes are often very correlated. WWR is especially common in CVAs when the default probability process and the interest rate process are positively correlated.

To examine the impact of this correlation on our agent's behavior, we train it on correlations $\rho \in \{-1, -0.5, 0, 0.5, 1\}$.

Trading cost

As additional market frictions make delta hedging non-optimal, we explore adding trading cost proportional to the value traded. Depending on the contract that the CVA comes from and market conditions, this is generally between 0 and 2 percent of the value of the trade. However, we choose to explore a slightly higher cost at 5% of the trade value and two unrealistic versions at exactly 0% and 100%. This corresponds to setting c to 0.05, 0, and 1 respectively in (3.2). The values were selected to test the ability for the reinforcement learning models to adapt to market frictions.

PPO parameters

Most hyperparameters use the default values from the library StableBaselines3. Table 8.1 shows the exact values we used. We try two network architectures for each environment, which we call the base and the deep models. Their respective sizes for environment 1 are shown in in Table 8.2. The base architecture is chosen based on performing well on standard RL test problems, which have a similar number of inputs and outputs. The deep one is then designed based on that one, such that it will have a similar number of parameters, as the number of parameters of a fully connected neural network scales as $\mathscr{O}(\text{width} * \text{depth}^2)$, where width is the number of nodes in a hidden layer and depth is the number of hidden layers. The deeper architecture was developed to test whether it could capture the non-linearities in our environment better. Table 8.3 shows the network architectures used for environment 2. Note that in environment 2, greater width is needed as the problem is of much higher dimension than in environment 1. Therefore, due to time constraints, the deep architecture here is much shallower than the deep architecture for environment 1, and thus also has much fewer parameters than the base for this environment. For environment 1, the base models are trained for 900 epochs while other models are trained for 700 epochs. For environment 2, both the base and deep models are trained for 900 epochs. After having been trained, the model that during training performed the best is saved and used. When evaluating, deterministic actions were used, meaning that the action with the highest density in the policy function was the action taken, instead of picking one according to the probability distribution.

Table 8.1 Hyperparameters for PPO

Parameters	Description	Value
η	Learning rate SGD	3×10^{-4}
n_{steps}	Number of steps per environment per update	2064
M	Mini batch size SGD	64
n_{epochs}	Number of epochs when optimizing the loss	10
γ	Discount factor	0.99
λ	GAE trade-off bias vs variance	0.95
ε	Clipping parameter, PPO loss clip range	0.2
c_1	Value function coefficient for loss calculation	0.5
c_2	Entropy coefficient for loss calculation	0.0

Table 8.2 Neural net architecture for PPO (environment 1) (both actor & critic have the same architecture)

Name	Size(# of neurons per layer times # of layers)
base	$[64 \times 2]$
deep	$[16 \times 32]$

Table 8.3 Neural net architecture for PPO (environment 2) (both actor & critic have the same architecture)

Name	Size(# of neurons per layer times # of layers)
base	$[512 \times 2]$
deep	$[64 \times 8]$

Pre-trained models

In environment 1 we also test pre-trained models for the market scenarios with trading costs. These models use the base neural network architecture and begins with the parameters from a model trained on no correlation and no trading cost with reward function 1 to begin their training.

8.2 Results

In this section we present our results, highlighting the most significant ones, and we will repeatedly refer to their corresponding tables in the appendix, marking the best-performing strategy with bold style on each metric.

Environment 1

Figure 11.4 shows a prototypical example of the learning curves for our base and deep models when the correlation and trading costs are both zero. One can see that they all improve and converge to that improved behavior. One can also see that reward 2 has a much higher variance than the others. Figure 11.3 is, on the other hand, an example of how the learning curves look when there is a very strong correlation, in this case $\rho = -1$, even with no trading cost. Here, all the models have higher variance; several do not improve significantly from their initial behaviors. Figure 11.5 shows that the results are similar when trading costs exist, even when ρ is zero.

Base model. One can see in Tables 11.1 – 11.5 that with no trading cost, the delta hedging performs the best on most metrics and correlation setups. One can also notice that the RL agents generally perform worse in extreme correlation set-ups in all metrics; see Tables 11.1 and 11.5. The models trained on Reward 2 and Reward 3 manage to make the most profits in some settings (see Tables 11.1, 11.2, and 11.4). The models trained on Reward 1 and Reward 3 come by far the closest to delta hedging, seen in Tables 11.2, 11.3, and 11.4. An example of how these good, trained, agents trade is shown in figure 11.2. One can see there that the jump in the default process has a significant negative effect.

For 5% trading cost, the delta hedging performs best on most metrics and correlation setups. But we notice that the RL agents perform worse in all set-ups, in all metrics, compared to without the trading cost, shown in Tables 11.6 – 11.10. Once again, some models make the most profits in all the settings. The RL agents trade much less compared to the no-trading cost setup.

Tables 11.11 – 11.15 show that for 100% trading cost, the strategy of "nothing" performs the best on most metrics and correlation setups. They also show that the RL agents learn to do almost no trading compared to before. Here, it is also evident that Delta becomes costly and drives the losses to higher levels compared to no and 5% trading cost.

Deep model. As with the base model, Tables 11.16 – 11.20 show that with no trading cost delta hedging is the best. Note however that they are never as close to delta hedging as the best models with the base architecture.

Tables 11.21 – 11.25 once again show that the reinforcement learning agents learn to hedge less when trading costs appear. However, they adjust too much, so they are still significantly worse than delta hedging for all metrics except mean losses.

When trading costs increase to 100%, however, we see new results. Table 11.26 still shows that with strong negative correlation and high trading costs. Looking at Tables 11.27 – 11.30 one can see though that the models are now often best at most metrics.

Pre-trained model. The learning curves for the pre-trained model look quite different from the ones for the previous models. An example is shown in Figure 11.6. The pattern is the same for all combinations of correlation and trading cost; they slowly worsen in performance. Looking at the results in Tables 11.31 – 11.40, one can see that there is never any single model which is best on everything, but that the trained models are consistently better than both delta hedging and the "nothing" strategy at several of the performance metrics.

Environment 2

Figure 11.10 shows the learning curves for our base models with reward function 1 when the trading costs are zero. One can see that it seems to converge to somewhere around -0.5 for all the correlations, with all of them except the one with -1 correlation ending with better behavior than they began. One can also see that it is much slower to converge for correlation 1. Figure 11.11 shows the learning curves for our base models with reward function 1 with 100% trading costs. Here one can see that they all converge to values significantly lower than they begin, except for the one with correlation -0.5, but it also has approximately the same value; so it not becoming worse simply shows that it began worse than the others.

Base reward 1. In Tables 11.41, 11.42, 11.43, and 11.45, the model is worse than both delta hedging and "nothing". In Table 11.44, it is better than "nothing" but still worse than delta hedging. With the low trading costs in Tables 11.48 and 11.50, the model is worse than both delta hedging and "nothing"; although in the same market conditions Tables 11.47 and 11.49 show that the model is better than "nothing" but still worse than delta hedging. Finally, in Table 11.46, it is better than "nothing" and delta hedging in several of the metrics. In the high trading cost situations shown in Tables 11.51 – 11.55, the model is always better than delta hedging on almost all metrics, and sometimes better on all the metrics. In Tables 11.52 – 11.55, it is also better than "nothing" on some but not all metrics. En example of how the hedge looks for the model presented in Table 11.52 is shown in Figure 11.8. One can see that it generally trades quite well, though there us a large loss in value at the one year mark where it held a lot of the swaption whose value became zero at that point. Another example of how the hedge looks for the same model is shown in Figure 11.9. One can see there once again that at the 4 year mark it held too much of a swaption that became worthless, but also that when a big jump happens in the default process at year 7, the model has no robustness, completely losing track of the process.

Deep reward 1. With no trading costs, Tables 11.56 – 11.60 show that the model is worse than both delta hedging and "nothing". Introducing low trading costs, Tables 11.61 – 11.63 continue to show that the model is worse than both delta hedging and "nothing". The model is better than "nothing" in Tables 11.64 and 11.65 though. In Tables 11.66 – 11.70, with high trading costs, the model is once again worse than both delta hedging and "nothing".

Base reward 7. In Tables 11.71, 11.72, 11.74, and 11.75, the model is worse than both delta hedging and "nothing" with no trading costs. In Table 11.73, it is better than "nothing" but still worse than delta hedging. Tables 11.76 – 11.79, all show that the model is worse than both delta hedging and "nothing" even with low trading costs. In Table 11.80, it is better than "nothing" but still worse than delta hedging. With very high trading costs, Tables 11.81, 11.82, 11.85, and 11.84 show the model performs better than delta on many performance metrics, but "nothing" is doing the best. Table 11.85 however shows delta doing better than our model, though the "nothing" strategy still outperforms delta.

Deep reward 7. One can see in Tables 11.86 – 11.90 that our models are always worse than both "nothing" and delta hedging with no trading costs. In Tables 11.91, 11.92, 11.93, and 11.95 one can see that our models are generally worse than both "nothing" and delta hedging even with low trading costs. In Table 11.94 the model outperforms the "nothing" strategy but is still worse than delta. Tables 11.96 – 11.99, and 11.100 that our models are once again always worse than both "nothing" and delta hedging, even with very high trading costs.

9

Discussion

This chapter will discuss the results from Chapter 8. It will discuss how well the agents performed compared to the other strategies, the effect of architecture choice, how pre-training a model improved the performance, different reward functions, how the correlations affected the performance, how RL agents tackled the non-linear effect of transaction costs, and learning convergence.

9.1 Overall performance

The observed trading behavior of the RL agent, as depicted in Figure11.2, suggests that the agent approximates delta hedging strategies under conditions of zero trading costs and correlation. In such a scenario, a delta hedging strategy would maintain constant values of one for both curves. However, the RL strategy underperforms compared to delta hedging in risk metrics due to the statistical improbability of policy distributions aligning precisely with right hedge values. Conversely, delta hedging does not account for market frictions like transaction costs, which opens opportunities for statistical learning approaches. These approaches can identify appropriate hedging strategies that require less frequent rebalancing, aiming for the right direction to offset P&L, while minimizing losses due to market frictions.

9.2 Environment 1

When trading costs are absent, delta hedging consistently emerges as the superior strategy. However, the introduction of trading costs yields more nuanced outcomes. Merely incorporating these costs relying on the base architecture proves ineffective. Instead, enhancing performance necessitates training more complex models or utilizing pre-trained models. These approaches can outperform both delta hedging and a passive strategy. Notably, when underlying processes are negatively correlated, they exhibit self-hedging properties, often resulting in the passive "do nothing" strategy performing surprisingly well.

9.3 Environment 2

Like in the first environment, delta hedging excels when there is no trading cost, and the "do nothing" strategy performs well under strong negative correlation. However, in this more complex environment, our RL models face greater challenges due to the increased dimension of the problem. Instead of determining two hedge quantities, agents must now identify twenty. Despite these challenges, it's noteworthy that our agents surpass the "do nothing" strategy in scenarios where it's crucial to act, particularly when there is a positive correlation. The fact that our models perform best at trading cost 1, which is wholly unrealistic, is not as useless as it might seem; as what it shows is that when the problem becomes significantly more non-linear, the RL models become better, and there are several complicating factors in real markets that we have not taken into account here. Figures 11.8 and 11.9 show that when the CVA behaves nicely, the hedge is able to follow it quite well, however when it jumps the RL models are unable to correctly respond. Also visible in those figures is how the agents don't seem to predict that swaptions will be worthless after their expiry, so one can see that the hedge value takes quite a large dip at the beginning of several years.

9.4 Neural network architecture

In scenarios without trading costs, delta hedging consistently outperforms other strategies across various metrics within the default architecture. However, when trading costs are introduced, their efficacy diminishes; although the variance in profit and loss P&L remains acceptable, other risk metrics show significant degradation. This performance drop is mainly due to increased losses driven by trading costs. In environments with trading costs, RL agents modify their approach by reducing trading frequency. Yet, despite these adjustments, the base architecture implementation of these agents does not achieve the same level of performance as delta hedging in any tested scenario of correlation and trading costs.

Modifying the neural network architecture of the actor and critic to a deeper configuration led to noticeable performance changes under high trading costs. This improvement is attributed to the deeper network's enhanced ability to approximate more complex, non-linear functions, which become increasingly critical as trading costs rise. However, in scenarios with no trading costs, delta hedging remains superior. In terms of broader risk metrics, delta hedging consistently outperforms, closely followed by RL agents, which do better than the "do nothing" strategy. While other strategies may also generate profits, they generally fall short in terms of variance and risk metrics.

Increasing the trading cost to 5% does not drastically change these outcomes, yet RL strategies begin to outperform both delta hedging and the "do nothing" strategy on several setups and metrics. When trading costs are raised to 100%, there is a significant shift in performance dynamics, RL strategies tend to outperform both delta hedging and the "do nothing" strategy across most metrics. The notable exception occurs when the correlation is -1, under which conditions the processes sufficiently hedge themselves, making the "do nothing" strategy the most effective.

9.5 Pre-trained models

Pre-trained models outperform other evaluated strategies at a 5% trading cost and align with the best deep models at a 100% trading cost. These models benefit from starting the learning process with an effective strategy, though this requires roughly twice the training time. Despite initial improvements, these models do not progress towards better performance, as evidenced by the instability of the optimum shown in Figure 11.6. Consequently, it is understandable why no single reward function stands out; all begin with equivalent behaviors, and any early improvements during the training phase are largely attributable to random explorations of network parameters rather than systematic advancements.

9.6 Reward functions

Rewards 1 and 3 generally perform well, but they encounter challenges in achieving consistent, positive behaviors under conditions of extreme correlation. Excluding these extreme cases, Rewards 1 and 3 often appear very similar and tend to outperform other rewards, likely because their simplicity facilitates easier learning by the networks.

Reward 2 occasionally manages to maximize profits, as designed. However, its effectiveness varies, and it is associated with significantly increased variance. This aligns with the financial models' principle that risk-free profits are impossible; thus, any profitable strategy must inherently assume greater risks, resulting in higher variance in outcomes.

Reward 5 is notable for its resilience to different levels of correlation, although it never leads in performance. Its robustness stems from its reward structure, which relies not only on actual outcomes but also on simulations of multiple potential scenarios. However, this makes it model-dependent and less ideal for real-world applications.

Rewards 4, 6, and 7 are commonly employed across diverse applications. Specifically, Reward 4 is a widely recognized mathematical metric, whereas Rewards 6 and 7 are popular risk metrics within the financial sector. These functions are used to penalize significant losses in the tail end of the distribution. Despite their prevalence, these rewards do not demonstrate superior performance compared to other rewards in the evaluations. Generally different reward functions affect learning, convergence, and performance. Suitable reward function engineering helps get better results in RL tasks.

9.7 Correlation

In environment 1, the extreme correlations of -1 and 1 caused the models to have difficulty learning. The nothing strategy performs best with negative correlation, as the processes hedge themselves there. Similarly, all strategies perform worse with positive correlation, as the opposite effect happens. These effects are most notable in environment 2, as intra-pair correlations and inter-pair correlations are present, where CVA is a weighted sum of Swaptions with weights as default probabilities.

9.8 Transaction costs

Across all setups, the introduction of trading costs leads models to reduce their trading amount, with higher trading costs prompting even greater reductions. This is the desired behavior, aligning perfectly with our expectations, as larger losses from costs imply larger penalties. However, without appropriate guidance, the models may reduce trading excessively, which can negatively impact the effective allocation of the correct hedge.

Incorporating guidance by initiating training with a pre-trained model addresses several challenges. This approach not only aids the agents in identifying effective strategies under low trading costs but also improves behavior under conditions of extreme correlation. Furthermore, employing a deeper network enables the capture of increased non-linearities associated with very high transaction costs, enhancing performance in such scenarios

9.9 Convergence

One of the main issues we observe in the learning of the RL agents is converging to a bad local optimum, which is seen clearly in Figures 11.5 and 11.11. One of the contributing factors to this might be the exploration and exploitation relation in PPO (discussed in 7). A further enhancement could be relaxing the exploitation nature of the algorithm, and instead encouraging exploration. These can be done by tuning the hyperparameters for the PPO algorithm in Table 8.1. Furthermore, the convergence of RL is significantly shaped by its initial conditions and specific aspects of its training regimen. This sensitivity stems from the ongoing collection of online data and the dependence on a singular scalar reward for feedback. Experiencing favorable training scenarios can greatly expedite the enhancement of a policy compared to others that may not encounter such beneficial conditions.

10

Conclusions

In this chapter we will discuss the conclusions that can be drawn from our results and their implications. In addition, we will present possible directions of future work in the subject.

10.1 Conclusions

Banks manage large trading books with exposure to various financial risks. To avoid scenarios like the 2008 financial crisis, oversight institutions have introduced laws and regulations to manage these risks. One such set of regulations is the *third Basel accords*, which crucially introduced the CVA [Bank for International Settlements, 2020]. CVA has been a significant component of P&L performance for Nordea and, more broadly, any large global financial institution offering derivatives. Research has shown that RL-based hedging strategies can outperform traditional methods when the contracts being hedged are quite simple, containing only one underlying asset and following simple market dynamics, even under conditions of transaction costs [Kolm and Ritter, 2019]. These advanced strategies leverage reinforcement learning to adapt to dynamic market conditions, providing a more robust framework for managing financial risks. The goal of this thesis has been to develop strategies to effectively manage the risks associated with CVA using similar RL-based methods, with the point being that CVA contains multiple underlying assets interacting in non-linear ways. Specifically, our research objectives have been

i) the development of deep hedging to manage the risks associated with CVA in simulated environments,

ii) the evaluation of deep hedging against a baseline method called *delta hedging*, and

iii) evaluation of the performance of the methods with respect to trading costs and underlying correlations in the simulated market environments.

To achieve these objectives we first implemented two market simulators, simulating CVA at different levels of complexity. We then trained PPO-based RL models to trade on the simulations, with varying reward functions, and varying correlations between the market forces and varying trading costs. To evaluate our models we implemented delta hedging and a nothing strategy. We then had our models and these strategies hedge the same realizations of the market simulations and compare their performance on several performance metrics.

We found that with no transaction costs, delta hedging always outperforms the RL-based models, although they can get very close. When transaction costs are added, however, the RL models can beat delta hedging on some metrics. Correlation has no clear impact on performance given that the models find reasonable behavior, however, it does have a large impact on the ability of the models to converge to something reasonable when it is large. This and the fact that the models have a hard time converging to reasonable behavior when trading costs are added indicates that optimal behavior is very unstable. It seems that the stochastic nature of these environments makes it challenging to learn the optimal strategy. Especially as the dimension of the problem increases.

10.2 Future work

This report explores a group of techniques that could enhance the performance of risk management proxies. The models showcased in this project are only advantageous with the existence of market frictions, which may indicate better performance in more realistic settings rather than theoretical frameworks. This showcases the great advantage of incorporating additional market information. While our findings provide valuable insights, they are based on simulated data. Conducting empirical validation with real-world financial data would be a crucial next step to confirm the applicability and effectiveness of the proposed strategies. We suggest here several factors to take into account for the models to be more realistic. We present also some suggestions to improve the performance of RL methods.

Model and environment development

Calibrating the model parameters to market data in both environments would enhance the simulation's alignment with real-world conditions. Such calibration would ensure that the strategies developed within this study are directly applicable to actual market actions, allowing for their evaluation in a more realistic context. This approach not only improves the reliability of the model outputs but also strengthens the practical relevance of the research findings. Incorporating CDS and swaps as hedging instruments, as how it is done at a CVA desk, rather than default probabilities and swaptions, would enhance the alignment of this study with prevailing industry practices, thus moving beyond our initial theoretical simplifications.

Furthermore, it is critical to consider the liquidity risk associated with these instruments, as their liquidity and availability can be sensitive to market fluctuations. Addressing these aspects will provide a more nuanced and realistic evaluation of the hedging strategies, thereby improving the model's applicability and robustness in practical financial environments.

An essential aspect of managing CVA is addressing both wrong-way and right-way risks. In this study, we have modeled scenarios where the correlation between exposure and default risk remains constant. However, in more realistic financial settings, these correlations may vary over time, introducing additional complexity to the problem. Therefore, it is advisable to employ robust, adaptive strategies capable of recognizing and responding to these dynamic risks effectively. This approach enhances the model's applicability and resilience in predicting realistic market behaviors.

Extending the development of the models and hedging strategies from individual contracts to encompass portfolios with netting would considerably enhance the realism and applicability of this study. This expansion would allow for a more sophisticated examination of interrelated risks and their mitigations within a portfolio context, providing insights that are more aligned with practical financial management and risk assessment practices

RL performance

In this study, the RL parameters were mostly the default settings provided by Sable-Baselines3, however it is well known that the selection of hyperparameters in RL significantly influences the performance outcomes of the agents. Considering hyperparameter optimization should improve the performance of RL agents significantly. As this aspect has been disregarded in this study it leaves a potential for improvement of RL strategies.

We also noted that trying deeper architectures for the agents improved the performance in certain situations. Research has shown that deeper architectures are better at handling non-linearities, and non-linearities are crucial aspects that RL methods could deliver on compared to other classical methods. Therefore, optimizing these deep architectures could advance the performance of RL agents. Moreover, pretrained models showed better results than trained from-scratch models. It might be helpful to build upon your best existing model to include additional complexities.

11

Appendix

11.1 Environment 1 tables

Base model

Table 11.1 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.96 \cdot 10^{-4}$	$1.97 \cdot 10^{-4}$	$1.97 \cdot 10^{-4}$	$-2.07 \cdot 10^{-4}$	$2.08 \cdot 10^{-4}$	$1.58 \cdot 10^{-2}$	$2.25 \cdot 10^{1}$
Reward 2	$4.31 \cdot 10^{-5}$	$4.33 \cdot 10^{-5}$	$4.33 \cdot 10^{-5}$	$-9.69 \cdot 10^{-5}$	$3.64 \cdot 10^{-4}$	$7.24 \cdot 10^{-3}$	$1.44 \cdot 10^{2}$
Reward 3	$7.18 \cdot 10^{-4}$	$7.23 \cdot 10^{-4}$	$7.23 \cdot 10^{-4}$	$\mathbf{-4.07 \cdot 10^{-4}}$	$3.38 \cdot 10^{-4}$	$2.96 \cdot 10^{-2}$	$5.23 \cdot 10^{1}$
Reward 4	$4.35 \cdot 10^{-5}$	$4.38 \cdot 10^{-5}$	$4.38 \cdot 10^{-5}$	$-9.55 \cdot 10^{-5}$	$4.21 \cdot 10^{-4}$	$7.33 \cdot 10^{-3}$	$3.81 \cdot 10^{2}$
Reward 5	$8.01 \cdot 10^{-7}$	$8.05 \cdot 10^{-7}$	$8.05 \cdot 10^{-7}$	$-5.61 \cdot 10^{-6}$	$3.56 \cdot 10^{-4}$	$9.80 \cdot 10^{-4}$	$1.09 \cdot 10^{3}$
Reward 6	$4.70 \cdot 10^{-6}$	$4.73 \cdot 10^{-6}$	$4.73 \cdot 10^{-6}$	$-2.89 \cdot 10^{-5}$	$4.88 \cdot 10^{-4}$	$2.31 \cdot 10^{-3}$	$4.61 \cdot 10^{3}$
Reward 7	$2.20 \cdot 10^{-5}$	$2.21 \cdot 10^{-5}$	$2.21 \cdot 10^{-5}$	$-6.56 \cdot 10^{-5}$	$4.42 \cdot 10^{-4}$	$5.26 \cdot 10^{-3}$	$1.64 \cdot 10^{2}$
Delta	$\mathbf{4.29 \cdot 10^{-10}}$	$\mathbf{4.31 \cdot 10^{-10}}$	$\mathbf{4.38 \cdot 10^{-10}}$	$2.51 \cdot 10^{-6}$	$\mathbf{1.78 \cdot 10^{-5}}$	$\mathbf{3.08 \cdot 10^{-5}}$	$1.23 \cdot 10^{3}$
Nothing	$7.24 \cdot 10^{-6}$	$7.26 \cdot 10^{-6}$	$7.27 \cdot 10^{-6}$	$-4.57 \cdot 10^{-5}$	$1.38 \cdot 10^{-3}$	$3.72 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.2 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$5.18 \cdot 10^{-9}$	$5.21 \cdot 10^{-9}$	$5.21 \cdot 10^{-9}$	$2.19 \cdot 10^{-6}$	$5.01 \cdot 10^{-5}$	$1.03 \cdot 10^{-4}$	$1.19 \cdot 10^{3}$
Reward 2	$6.45 \cdot 10^{-5}$	$6.48 \cdot 10^{-5}$	$6.48 \cdot 10^{-5}$	$\mathbf{-1.65 \cdot 10^{-4}}$	$2.02 \cdot 10^{-3}$	$1.14 \cdot 10^{-2}$	$1.13 \cdot 10^{2}$
Reward 3	$1.00 \cdot 10^{-8}$	$1.01 \cdot 10^{-8}$	$1.01 \cdot 10^{-8}$	$1.87 \cdot 10^{-7}$	$3.97 \cdot 10^{-5}$	$1.05 \cdot 10^{-4}$	$1.18 \cdot 10^{3}$
Reward 4	$8.16 \cdot 10^{-5}$	$8.21 \cdot 10^{-5}$	$8.21 \cdot 10^{-5}$	$1.78 \cdot 10^{-4}$	$2.53 \cdot 10^{-3}$	$1.16 \cdot 10^{-2}$	$7.63 \cdot 10^{4}$
Reward 5	$2.91 \cdot 10^{-7}$	$2.91 \cdot 10^{-7}$	$2.91 \cdot 10^{-7}$	$2.24 \cdot 10^{-6}$	$6.02 \cdot 10^{-4}$	$9.79 \cdot 10^{-4}$	$8.25 \cdot 10^{2}$
Reward 6	$2.06 \cdot 10^{-7}$	$2.06 \cdot 10^{-7}$	$2.06 \cdot 10^{-7}$	$-3.44 \cdot 10^{-6}$	$3.17 \cdot 10^{-4}$	$6.76 \cdot 10^{-4}$	$1.11 \cdot 10^{3}$
Reward 7	$6.80 \cdot 10^{-8}$	$6.82 \cdot 10^{-8}$	$6.82 \cdot 10^{-8}$	$5.76 \cdot 10^{-7}$	$1.31 \cdot 10^{-4}$	$3.51 \cdot 10^{-4}$	$1.05 \cdot 10^{3}$
Delta	$\mathbf{7.10 \cdot 10^{-10}}$	$\mathbf{7.14 \cdot 10^{-10}}$	$\mathbf{7.16 \cdot 10^{-10}}$	$1.53 \cdot 10^{-6}$	$\mathbf{1.39 \cdot 10^{-5}}$	$\mathbf{3.21 \cdot 10^{-5}}$	$1.23 \cdot 10^{3}$
Nothing	$9.02 \cdot 10^{-6}$	$9.05 \cdot 10^{-6}$	$9.05 \cdot 10^{-6}$	$-6.86 \cdot 10^{-5}$	$1.94 \cdot 10^{-3}$	$5.08 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.3 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.21 \cdot 10^{-9}$	$2.21 \cdot 10^{-9}$	$2.21 \cdot 10^{-9}$	$-2.13 \cdot 10^{-7}$	$4.55 \cdot 10^{-5}$	$7.85 \cdot 10^{-5}$	$1.04 \cdot 10^{3}$
Reward 2	$2.30 \cdot 10^{-6}$	$2.30 \cdot 10^{-6}$	$2.30 \cdot 10^{-6}$	$-2.93 \cdot 10^{-5}$	$1.84 \cdot 10^{-3}$	$2.99 \cdot 10^{-3}$	$3.22 \cdot 10^{3}$
Reward 3	$1.19 \cdot 10^{-8}$	$1.19 \cdot 10^{-8}$	$1.19 \cdot 10^{-8}$	$6.85 \cdot 10^{-7}$	$7.14 \cdot 10^{-5}$	$1.69 \cdot 10^{-4}$	$1.00 \cdot 10^{3}$
Reward 4	$9.01 \cdot 10^{-6}$	$9.04 \cdot 10^{-6}$	$9.04 \cdot 10^{-6}$	$1.89 \cdot 10^{-5}$	$2.29 \cdot 10^{-3}$	$4.87 \cdot 10^{-3}$	$5.04 \cdot 10^{2}$
Reward 5	$4.59 \cdot 10^{-7}$	$4.60 \cdot 10^{-7}$	$4.60 \cdot 10^{-7}$	$-9.91 \cdot 10^{-7}$	$7.91 \cdot 10^{-4}$	$1.25 \cdot 10^{-3}$	$1.03 \cdot 10^{3}$
Reward 6	$5.40 \cdot 10^{-7}$	$5.41 \cdot 10^{-7}$	$5.42 \cdot 10^{-7}$	$-1.02 \cdot 10^{-5}$	$2.14 \cdot 10^{-4}$	$7.86 \cdot 10^{-4}$	$1.04 \cdot 10^{3}$
Reward 7	$1.10 \cdot 10^{-7}$	$1.11 \cdot 10^{-7}$	$1.11 \cdot 10^{-7}$	$3.15 \cdot 10^{-6}$	$1.17 \cdot 10^{-4}$	$4.20 \cdot 10^{-4}$	$9.46 \cdot 10^{2}$
Delta	$\mathbf{8.15 \cdot 10^{-10}}$	$\mathbf{8.16 \cdot 10^{-10}}$	$\mathbf{8.16 \cdot 10^{-10}}$	$-4.31 \cdot 10^{-8}$	$\mathbf{8.71 \cdot 10^{-6}}$	$\mathbf{3.46 \cdot 10^{-5}}$	$1.19 \cdot 10^{3}$
Nothing	$1.30 \cdot 10^{-5}$	$1.30 \cdot 10^{-5}$	$1.30 \cdot 10^{-5}$	$\mathbf{-7.17 \cdot 10^{-5}}$	$2.32 \cdot 10^{-3}$	$5.91 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.4 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.14 \cdot 10^{-8}$	$1.15 \cdot 10^{-8}$	$1.15 \cdot 10^{-8}$	$-2.49 \cdot 10^{-7}$	$6.73 \cdot 10^{-5}$	$1.48 \cdot 10^{-4}$	$1.19 \cdot 10^{3}$
Reward 2	$4.30 \cdot 10^{-5}$	$4.33 \cdot 10^{-5}$	$4.33 \cdot 10^{-5}$	$\mathbf{-1.24 \cdot 10^{-4}}$	$3.25 \cdot 10^{-3}$	$9.62 \cdot 10^{-3}$	$4.99 \cdot 10^{2}$
Reward 3	$7.71 \cdot 10^{-9}$	$7.71 \cdot 10^{-9}$	$7.72 \cdot 10^{-9}$	$-1.23 \cdot 10^{-6}$	$4.59 \cdot 10^{-5}$	$1.25 \cdot 10^{-4}$	$1.09 \cdot 10^{3}$
Reward 4	$1.89 \cdot 10^{-4}$	$1.91 \cdot 10^{-4}$	$1.91 \cdot 10^{-4}$	$2.33 \cdot 10^{-4}$	$6.39 \cdot 10^{-3}$	$1.91 \cdot 10^{-2}$	$5.68 \cdot 10^{2}$
Reward 5	$2.54 \cdot 10^{-7}$	$2.55 \cdot 10^{-7}$	$2.55 \cdot 10^{-7}$	$-6.96 \cdot 10^{-6}$	$2.20 \cdot 10^{-4}$	$7.34 \cdot 10^{-4}$	$1.07 \cdot 10^{3}$
Reward 6	$7.15 \cdot 10^{-7}$	$7.18 \cdot 10^{-7}$	$7.19 \cdot 10^{-7}$	$8.28 \cdot 10^{-6}$	$3.37 \cdot 10^{-4}$	$1.00 \cdot 10^{-3}$	$1.05 \cdot 10^{3}$
Reward 7	$8.84 \cdot 10^{-8}$	$8.85 \cdot 10^{-8}$	$8.85 \cdot 10^{-8}$	$1.08 \cdot 10^{-6}$	$1.36 \cdot 10^{-4}$	$3.54 \cdot 10^{-4}$	$1.08 \cdot 10^{3}$
Delta	$\mathbf{4.95 \cdot 10^{-10}}$	$\mathbf{4.99 \cdot 10^{-10}}$	$\mathbf{5.02 \cdot 10^{-10}}$	$-1.56 \cdot 10^{-6}$	$\mathbf{3.68 \cdot 10^{-6}}$	$\mathbf{2.89 \cdot 10^{-5}}$	$1.23 \cdot 10^{3}$
Nothing	$1.04 \cdot 10^{-5}$	$1.05 \cdot 10^{-5}$	$1.05 \cdot 10^{-5}$	$-6.52 \cdot 10^{-5}$	$2.48 \cdot 10^{-3}$	$5.49 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.5 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.38 \cdot 10^{-5}$	$1.38 \cdot 10^{-5}$	$1.38 \cdot 10^{-5}$	$2.81 \cdot 10^{-5}$	$1.34 \cdot 10^{-4}$	$5.05 \cdot 10^{-3}$	$3.31 \cdot 10^{2}$
Reward 2	$4.38 \cdot 10^{-4}$	$4.39 \cdot 10^{-4}$	$4.39 \cdot 10^{-4}$	$1.65 \cdot 10^{-4}$	$6.87 \cdot 10^{-4}$	$2.93 \cdot 10^{-2}$	$2.81 \cdot 10^{2}$
Reward 3	$2.51 \cdot 10^{-5}$	$2.51 \cdot 10^{-5}$	$2.51 \cdot 10^{-5}$	$3.84 \cdot 10^{-5}$	$1.26 \cdot 10^{-4}$	$6.81 \cdot 10^{-3}$	$3.11 \cdot 10^{2}$
Reward 4	$3.39 \cdot 10^{-6}$	$3.39 \cdot 10^{-6}$	$3.39 \cdot 10^{-6}$	$-1.06 \cdot 10^{-5}$	$7.10 \cdot 10^{-4}$	$2.52 \cdot 10^{-3}$	$3.54 \cdot 10^{2}$
Reward 5	$2.87 \cdot 10^{-7}$	$2.87 \cdot 10^{-7}$	$2.87 \cdot 10^{-7}$	$-6.46 \cdot 10^{-6}$	$2.33 \cdot 10^{-4}$	$8.93 \cdot 10^{-4}$	$1.38 \cdot 10^{3}$
Reward 6	$7.78 \cdot 10^{-6}$	$7.79 \cdot 10^{-6}$	$7.79 \cdot 10^{-6}$	$1.72 \cdot 10^{-5}$	$2.29 \cdot 10^{-4}$	$3.88 \cdot 10^{-3}$	$4.49 \cdot 10^{2}$
Reward 7	$2.05 \cdot 10^{-6}$	$2.05 \cdot 10^{-6}$	$2.05 \cdot 10^{-6}$	$6.75 \cdot 10^{-6}$	$2.94 \cdot 10^{-4}$	$2.06 \cdot 10^{-3}$	$3.88 \cdot 10^{2}$
Delta	$\mathbf{1.18 \cdot 10^{-9}}$	$\mathbf{1.18 \cdot 10^{-9}}$	$\mathbf{1.19 \cdot 10^{-9}}$	$-3.13 \cdot 10^{-6}$	$\mathbf{1.35 \cdot 10^{-9}}$	$\mathbf{3.29 \cdot 10^{-5}}$	$1.19 \cdot 10^{3}$
Nothing	$9.89 \cdot 10^{-6}$	$9.90 \cdot 10^{-6}$	$9.90 \cdot 10^{-6}$	$\mathbf{-5.61 \cdot 10^{-5}}$	$2.99 \cdot 10^{-3}$	$5.62 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.6 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.14\cdot10^{-3}$	$1.14\cdot10^{-3}$	$1.15\cdot10^{-3}$	$-6.98\cdot10^{-4}$	$3.89\cdot10^{-3}$	$3.93\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Reward 2	$1.14\cdot10^{-3}$	$1.14\cdot10^{-3}$	$1.15\cdot10^{-3}$	$\mathbf{-6.98\cdot10^{-4}}$	$3.89\cdot10^{-3}$	$3.93\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Reward 3	$1.13\cdot10^{-3}$	$1.14\cdot10^{-3}$	$1.14\cdot10^{-3}$	$-6.66\cdot10^{-4}$	$2.26\cdot10^{-3}$	$3.77\cdot10^{-2}$	$1.40\cdot10^{1}$
Reward 4	$8.20\cdot10^{-4}$	$8.26\cdot10^{-4}$	$8.29\cdot10^{-4}$	$6.90\cdot10^{-4}$	$1.12\cdot10^{-2}$	$3.38\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Reward 5	$1.14\cdot10^{-3}$	$1.14\cdot10^{-3}$	$1.15\cdot10^{-3}$	$\mathbf{-6.98\cdot10^{-4}}$	$3.89\cdot10^{-3}$	$3.93\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Reward 6	$8.21\cdot10^{-4}$	$8.26\cdot10^{-4}$	$8.30\cdot10^{-4}$	$6.90\cdot10^{-4}$	$1.13\cdot10^{-2}$	$3.39\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Reward 7	$8.21\cdot10^{-4}$	$8.26\cdot10^{-4}$	$8.30\cdot10^{-4}$	$6.90\cdot10^{-4}$	$1.13\cdot10^{-2}$	$3.39\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Delta	$\mathbf{5.82\cdot10^{-8}}$	$\mathbf{3.32\cdot10^{-9}}$	$\mathbf{6.22\cdot10^{-8}}$	$5.85\cdot10^{-5}$	$\mathbf{2.26\cdot10^{-4}}$	$\mathbf{5.84\cdot10^{-4}}$	$1.23\cdot10^{3}$
Nothing	$7.33\cdot10^{-6}$	$7.38\cdot10^{-6}$	$7.39\cdot10^{-6}$	$-6.33\cdot10^{-5}$	$1.50\cdot10^{-3}$	$4.07\cdot10^{-3}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.7 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.32\cdot10^{-3}$	$1.32\cdot10^{-3}$	$1.32\cdot10^{-3}$	$\mathbf{-4.03\cdot10^{-4}}$	$7.11\cdot10^{-3}$	$3.93\cdot10^{-2}$	$2.25\cdot10^{1}$
Reward 2	$7.68\cdot10^{-4}$	$7.71\cdot10^{-4}$	$7.71\cdot10^{-4}$	$3.49\cdot10^{-4}$	$7.51\cdot10^{-3}$	$3.20\cdot10^{-2}$	$3.33\cdot10^{1}$
Reward 3	$6.28\cdot10^{-4}$	$6.30\cdot10^{-4}$	$6.30\cdot10^{-4}$	$3.50\cdot10^{-4}$	$4.84\cdot10^{-3}$	$2.74\cdot10^{-2}$	$2.23\cdot10^{1}$
Reward 4	$1.07\cdot10^{-3}$	$1.07\cdot10^{-3}$	$1.07\cdot10^{-3}$	$-3.67\cdot10^{-4}$	$6.46\cdot10^{-3}$	$3.54\cdot10^{-2}$	$1.59\cdot10^{1}$
Reward 5	$8.39\cdot10^{-4}$	$8.42\cdot10^{-4}$	$8.41\cdot10^{-4}$	$-2.97\cdot10^{-4}$	$7.06\cdot10^{-3}$	$3.29\cdot10^{-2}$	$5.13\cdot10^{1}$
Reward 6	$2.25\cdot10^{-4}$	$2.26\cdot10^{-4}$	$2.26\cdot10^{-4}$	$-1.52\cdot10^{-4}$	$2.78\cdot10^{-3}$	$1.43\cdot10^{-2}$	$5.17\cdot10^{2}$
Reward 7	$1.03\cdot10^{-3}$	$1.04\cdot10^{-3}$	$1.04\cdot10^{-3}$	$4.88\cdot10^{-4}$	$7.60\cdot10^{-3}$	$3.62\cdot10^{-2}$	$7.02\cdot10^{0}$
Delta	$\mathbf{4.77\cdot10^{-8}}$	$\mathbf{3.65\cdot10^{-9}}$	$\mathbf{5.09\cdot10^{-8}}$	$5.46\cdot10^{-5}$	$\mathbf{1.50\cdot10^{-4}}$	$\mathbf{5.64\cdot10^{-4}}$	$1.17\cdot10^{3}$
Nothing	$1.09\cdot10^{-5}$	$1.10\cdot10^{-5}$	$1.10\cdot10^{-5}$	$-5.29\cdot10^{-5}$	$1.91\cdot10^{-3}$	$4.67\cdot10^{-3}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.8 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$6.98\cdot10^{-3}$	$7.52\cdot10^{-3}$	$7.61\cdot10^{-3}$	$2.71\cdot10^{-3}$	$3.16\cdot10^{-2}$	$4.99\cdot10^{-2}$	$2.62\cdot10^{1}$
Reward 2	$1.07\cdot10^{-2}$	$1.16\cdot10^{-2}$	$1.16\cdot10^{-2}$	$-3.21\cdot10^{-3}$	$9.32\cdot10^{-3}$	$5.78\cdot10^{-2}$	$1.53\cdot10^{1}$
Reward 3	$4.72\cdot10^{-3}$	$5.07\cdot10^{-3}$	$5.10\cdot10^{-3}$	$-2.11\cdot10^{-3}$	$5.91\cdot10^{-3}$	$4.29\cdot10^{-2}$	$1.87\cdot10^{1}$
Reward 4	$7.22\cdot10^{-3}$	$7.77\cdot10^{-3}$	$7.80\cdot10^{-3}$	$-2.60\cdot10^{-3}$	$6.52\cdot10^{-3}$	$4.97\cdot10^{-2}$	$2.08\cdot10^{1}$
Reward 5	$1.22\cdot10^{-2}$	$1.32\cdot10^{-2}$	$1.32\cdot10^{-2}$	$\mathbf{-3.37\cdot10^{-3}}$	$7.71\cdot10^{-3}$	$6.06\cdot10^{-2}$	$1.80\cdot10^{1}$
Reward 6	$4.68\cdot10^{-3}$	$5.04\cdot10^{-3}$	$5.10\cdot10^{-3}$	$2.23\cdot10^{-3}$	$2.59\cdot10^{-2}$	$4.17\cdot10^{-2}$	$5.50\cdot10^{1}$
Reward 7	$1.03\cdot10^{-2}$	$1.11\cdot10^{-2}$	$1.12\cdot10^{-2}$	$3.34\cdot10^{-3}$	$3.91\cdot10^{-2}$	$6.13\cdot10^{-2}$	$2.98\cdot10^{1}$
Delta	$\mathbf{1.22\cdot10^{-7}}$	$\mathbf{3.20\cdot10^{-9}}$	$\mathbf{1.35\cdot10^{-7}}$	$6.02\cdot10^{-5}$	$\mathbf{2.34\cdot10^{-4}}$	$\mathbf{5.67\cdot10^{-4}}$	$1.22\cdot10^{3}$
Nothing	$6.91\cdot10^{-5}$	$7.43\cdot10^{-5}$	$7.49\cdot10^{-5}$	$-2.78\cdot10^{-4}$	$2.24\cdot10^{-3}$	$6.29\cdot10^{-3}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.9 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$7.55\cdot10^{-4}$	$7.57\cdot10^{-4}$	$7.57\cdot10^{-4}$	$3.55\cdot10^{-4}$	$5.76\cdot10^{-3}$	$3.47\cdot10^{-2}$	$3.35\cdot10^{1}$
Reward 2	$7.57\cdot10^{-4}$	$7.59\cdot10^{-4}$	$7.59\cdot10^{-4}$	$\mathbf{-3.23\cdot10^{-4}}$	$8.06\cdot10^{-3}$	$3.59\cdot10^{-2}$	$2.73\cdot10^{1}$
Reward 3	$6.45\cdot10^{-4}$	$6.47\cdot10^{-4}$	$6.47\cdot10^{-4}$	$2.80\cdot10^{-4}$	$5.61\cdot10^{-3}$	$3.28\cdot10^{-2}$	$3.14\cdot10^{1}$
Reward 4	$5.87\cdot10^{-4}$	$5.88\cdot10^{-4}$	$5.88\cdot10^{-4}$	$-2.60\cdot10^{-4}$	$6.78\cdot10^{-3}$	$3.16\cdot10^{-2}$	$5.42\cdot10^{1}$
Reward 5	$7.25\cdot10^{-4}$	$7.27\cdot10^{-4}$	$7.27\cdot10^{-4}$	$-3.14\cdot10^{-4}$	$7.83\cdot10^{-3}$	$3.54\cdot10^{-2}$	$3.31\cdot10^{1}$
Reward 6	$2.22\cdot10^{-4}$	$2.22\cdot10^{-4}$	$2.22\cdot10^{-4}$	$2.33\cdot10^{-4}$	$4.72\cdot10^{-3}$	$2.01\cdot10^{-2}$	$5.31\cdot10^{1}$
Reward 7	$4.83\cdot10^{-4}$	$4.84\cdot10^{-4}$	$4.84\cdot10^{-4}$	$-2.87\cdot10^{-5}$	$5.45\cdot10^{-3}$	$2.72\cdot10^{-2}$	$4.63\cdot10^{1}$
Delta	$\mathbf{4.09\cdot10^{-8}}$	$\mathbf{2.98\cdot10^{-9}}$	$\mathbf{4.37\cdot10^{-8}}$	$4.99\cdot10^{-5}$	$\mathbf{1.45\cdot10^{-4}}$	$\mathbf{5.12\cdot10^{-4}}$	$1.28\cdot10^{3}$
Nothing	$8.70\cdot10^{-6}$	$8.71\cdot10^{-6}$	$8.72\cdot10^{-6}$	$-5.23\cdot10^{-5}$	$2.55\cdot10^{-3}$	$5.15\cdot10^{-3}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.10 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.15\cdot10^{-3}$	$1.16\cdot10^{-3}$	$1.16\cdot10^{-3}$	$-4.92\cdot10^{-4}$	$3.43\cdot10^{-3}$	$3.93\cdot10^{-2}$	$8.74\cdot10^{1}$
Reward 2	$1.15\cdot10^{-3}$	$1.16\cdot10^{-3}$	$1.16\cdot10^{-3}$	$-4.91\cdot10^{-4}$	$3.56\cdot10^{-3}$	$3.91\cdot10^{-2}$	$8.72\cdot10^{1}$
Reward 3	$1.15\cdot10^{-3}$	$1.16\cdot10^{-3}$	$1.16\cdot10^{-3}$	$-4.91\cdot10^{-4}$	$3.53\cdot10^{-3}$	$3.92\cdot10^{-2}$	$8.73\cdot10^{1}$
Reward 4	$1.15\cdot10^{-3}$	$1.16\cdot10^{-3}$	$1.16\cdot10^{-3}$	$-4.92\cdot10^{-4}$	$3.32\cdot10^{-3}$	$3.96\cdot10^{-2}$	$8.74\cdot10^{1}$
Reward 5	$1.15\cdot10^{-3}$	$1.16\cdot10^{-3}$	$1.16\cdot10^{-3}$	$\mathbf{-4.92\cdot10^{-4}}$	$3.47\cdot10^{-3}$	$3.92\cdot10^{-2}$	$8.75\cdot10^{1}$
Reward 6	$1.15\cdot10^{-3}$	$1.16\cdot10^{-3}$	$1.16\cdot10^{-3}$	$-4.92\cdot10^{-4}$	$3.32\cdot10^{-3}$	$3.96\cdot10^{-2}$	$8.74\cdot10^{1}$
Reward 7	$1.16\cdot10^{-3}$	$1.16\cdot10^{-3}$	$1.16\cdot10^{-3}$	$-4.19\cdot10^{-4}$	$4.13\cdot10^{-3}$	$4.06\cdot10^{-2}$	$2.70\cdot10^{2}$
Delta	$\mathbf{5.07\cdot10^{-8}}$	$\mathbf{3.36\cdot10^{-9}}$	$\mathbf{5.36\cdot10^{-8}}$	$5.08\cdot10^{-5}$	$\mathbf{1.96\cdot10^{-4}}$	$\mathbf{5.57\cdot10^{-4}}$	$1.15\cdot10^{3}$
Nothing	$8.22\cdot10^{-6}$	$8.27\cdot10^{-6}$	$8.27\cdot10^{-6}$	$-6.37\cdot10^{-5}$	$2.85\cdot10^{-3}$	$5.31\cdot10^{-3}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.11 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$3.20\cdot10^{-3}$	$1.72\cdot10^{-3}$	$3.20\cdot10^{-3}$	$2.04\cdot10^{-4}$	$1.08\cdot10^{-2}$	$1.08\cdot10^{-1}$	$\mathbf{0.00\cdot10^{0}}$
Reward 2	$1.88\cdot10^{-3}$	$1.84\cdot10^{-3}$	$1.91\cdot10^{-3}$	$1.75\cdot10^{-3}$	$2.08\cdot10^{-2}$	$7.99\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Reward 3	$3.20\cdot10^{-3}$	$1.72\cdot10^{-3}$	$3.20\cdot10^{-3}$	$2.04\cdot10^{-4}$	$1.08\cdot10^{-2}$	$1.08\cdot10^{-1}$	$\mathbf{0.00\cdot10^{0}}$
Reward 4	$2.67\cdot10^{-3}$	$1.23\cdot10^{-3}$	$2.71\cdot10^{-3}$	$1.75\cdot10^{-3}$	$2.52\cdot10^{-2}$	$1.03\cdot10^{-1}$	$\mathbf{0.00\cdot10^{0}}$
Reward 5	$2.43\cdot10^{-3}$	$1.24\cdot10^{-3}$	$2.47\cdot10^{-3}$	$1.75\cdot10^{-3}$	$2.38\cdot10^{-2}$	$9.65\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Reward 6	$3.21\cdot10^{-3}$	$1.76\cdot10^{-3}$	$3.21\cdot10^{-3}$	$3.74\cdot10^{-4}$	$2.18\cdot10^{-2}$	$1.08\cdot10^{-1}$	$8.85\cdot10^{1}$
Reward 7	$2.60\cdot10^{-3}$	$1.23\cdot10^{-3}$	$2.64\cdot10^{-3}$	$1.75\cdot10^{-3}$	$2.45\cdot10^{-2}$	$1.00\cdot10^{-1}$	$\mathbf{0.00\cdot10^{0}}$
Delta	$2.45\cdot10^{-5}$	$\mathbf{1.35\cdot10^{-6}}$	$2.59\cdot10^{-5}$	$1.08\cdot10^{-3}$	$4.14\cdot10^{-3}$	$1.14\cdot10^{-2}$	$1.16\cdot10^{3}$
Nothing	$\mathbf{1.14\cdot10^{-5}}$	$1.15\cdot10^{-5}$	$\mathbf{1.15\cdot10^{-5}}$	$\mathbf{-7.88\cdot10^{-5}}$	$\mathbf{1.43\cdot10^{-3}}$	$\mathbf{5.05\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.12 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$9.71\cdot10^{-4}$	$9.71\cdot10^{-4}$	$9.79\cdot10^{-4}$	$1.08\cdot10^{-3}$	$2.03\cdot10^{-2}$	$5.20\cdot10^{-2}$	$6.74\cdot10^{0}$
Reward 2	$2.00\cdot10^{-3}$	$8.47\cdot10^{-4}$	$2.01\cdot10^{-3}$	$1.32\cdot10^{-3}$	$1.53\cdot10^{-2}$	$9.55\cdot10^{-2}$	$2.29\cdot10^{-2}$
Reward 3	$5.16\cdot10^{-4}$	$5.17\cdot10^{-4}$	$5.19\cdot10^{-4}$	$7.91\cdot10^{-4}$	$1.28\cdot10^{-2}$	$3.30\cdot10^{-2}$	$4.41\cdot10^{1}$
Reward 4	$8.14\cdot10^{-4}$	$8.17\cdot10^{-4}$	$8.16\cdot10^{-4}$	$3.93\cdot10^{-4}$	$1.32\cdot10^{-2}$	$4.15\cdot10^{-2}$	$1.67\cdot10^{1}$
Reward 5	$1.26\cdot10^{-3}$	$1.27\cdot10^{-3}$	$1.27\cdot10^{-3}$	$1.95\cdot10^{-4}$	$1.46\cdot10^{-2}$	$5.07\cdot10^{-2}$	$1.15\cdot10^{1}$
Reward 6	$7.81\cdot10^{-4}$	$7.75\cdot10^{-4}$	$7.82\cdot10^{-4}$	$3.39\cdot10^{-4}$	$1.32\cdot10^{-2}$	$4.10\cdot10^{-2}$	$4.99\cdot10^{1}$
Reward 7	$6.82\cdot10^{-4}$	$6.76\cdot10^{-4}$	$6.83\cdot10^{-4}$	$2.29\cdot10^{-4}$	$1.08\cdot10^{-2}$	$3.75\cdot10^{-2}$	$2.54\cdot10^{1}$
Delta	$2.24\cdot10^{-5}$	$\mathbf{1.10\cdot10^{-6}}$	$2.36\cdot10^{-5}$	$1.06\cdot10^{-3}$	$2.93\cdot10^{-3}$	$1.10\cdot10^{-2}$	$1.35\cdot10^{3}$
Nothing	$\mathbf{1.36\cdot10^{-5}}$	$1.36\cdot10^{-5}$	$\mathbf{1.36\cdot10^{-5}}$	$\mathbf{-6.35\cdot10^{-5}}$	$\mathbf{1.91\cdot10^{-3}}$	$\mathbf{5.07\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.13 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$7.16\cdot10^{-4}$	$7.25\cdot10^{-4}$	$7.52\cdot10^{-4}$	$1.34\cdot10^{-3}$	$1.96\cdot10^{-2}$	$4.35\cdot10^{-2}$	$1.09\cdot10^{1}$
Reward 2	$2.92\cdot10^{-5}$	$2.83\cdot10^{-5}$	$2.93\cdot10^{-5}$	$2.56\cdot10^{-4}$	$3.27\cdot10^{-3}$	$8.19\cdot10^{-3}$	$8.33\cdot10^{2}$
Reward 3	$2.39\cdot10^{-3}$	$9.55\cdot10^{-4}$	$2.39\cdot10^{-3}$	$4.23\cdot10^{-4}$	$2.29\cdot10^{-2}$	$9.41\cdot10^{-2}$	$\mathbf{0.00\cdot10^{0}}$
Reward 4	$6.16\cdot10^{-4}$	$6.24\cdot10^{-4}$	$6.49\cdot10^{-4}$	$1.29\cdot10^{-3}$	$1.96\cdot10^{-2}$	$3.87\cdot10^{-2}$	$1.49\cdot10^{1}$
Reward 5	$4.15\cdot10^{-4}$	$4.13\cdot10^{-4}$	$4.37\cdot10^{-4}$	$1.04\cdot10^{-3}$	$1.49\cdot10^{-2}$	$3.30\cdot10^{-2}$	$1.45\cdot10^{1}$
Reward 6	$4.87\cdot10^{-4}$	$4.93\cdot10^{-4}$	$4.88\cdot10^{-4}$	$2.55\cdot10^{-4}$	$1.10\cdot10^{-2}$	$3.11\cdot10^{-2}$	$1.90\cdot10^{1}$
Reward 7	$6.99\cdot10^{-4}$	$7.09\cdot10^{-4}$	$7.38\cdot10^{-4}$	$1.38\cdot10^{-3}$	$1.96\cdot10^{-2}$	$4.25\cdot10^{-2}$	$5.78\cdot10^{0}$
Delta	$2.51\cdot10^{-5}$	$\mathbf{1.18\cdot10^{-6}}$	$2.66\cdot10^{-5}$	$1.06\cdot10^{-3}$	$4.05\cdot10^{-3}$	$1.08\cdot10^{-2}$	$1.19\cdot10^{3}$
Nothing	$\mathbf{8.40\cdot10^{-6}}$	$8.52\cdot10^{-6}$	$\mathbf{8.52\cdot10^{-6}}$	$\mathbf{-6.65\cdot10^{-5}}$	$\mathbf{2.21\cdot10^{-3}}$	$\mathbf{4.54\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.14 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.40\cdot10^{-3}$	$1.48\cdot10^{-3}$	$1.40\cdot10^{-3}$	$3.66\cdot10^{-4}$	$9.14\cdot10^{-3}$	$3.42\cdot10^{-2}$	$1.18\cdot10^{2}$
Reward 2	$1.77\cdot10^{-3}$	$1.87\cdot10^{-3}$	$1.79\cdot10^{-3}$	$-1.78\cdot10^{-4}$	$1.09\cdot10^{-2}$	$3.85\cdot10^{-2}$	$1.49\cdot10^{1}$
Reward 3	$1.60\cdot10^{-3}$	$1.68\cdot10^{-3}$	$1.96\cdot10^{-3}$	$2.62\cdot10^{-3}$	$2.14\cdot10^{-2}$	$4.41\cdot10^{-2}$	$4.00\cdot10^{1}$
Reward 4	$6.20\cdot10^{-3}$	$3.43\cdot10^{-3}$	$6.21\cdot10^{-3}$	$2.29\cdot10^{-4}$	$2.20\cdot10^{-2}$	$1.04\cdot10^{-1}$	$3.44\cdot10^{2}$
Reward 5	$1.34\cdot10^{-3}$	$1.41\cdot10^{-3}$	$1.65\cdot10^{-3}$	$2.42\cdot10^{-3}$	$2.03\cdot10^{-2}$	$4.09\cdot10^{-2}$	$1.39\cdot10^{1}$
Reward 6	$1.42\cdot10^{-3}$	$1.50\cdot10^{-3}$	$1.76\cdot10^{-3}$	$2.52\cdot10^{-3}$	$1.95\cdot10^{-2}$	$4.28\cdot10^{-2}$	$3.26\cdot10^{1}$
Reward 7	$2.08\cdot10^{-4}$	$2.14\cdot10^{-4}$	$2.09\cdot10^{-4}$	$2.93\cdot10^{-4}$	$9.28\cdot10^{-3}$	$1.79\cdot10^{-2}$	$1.39\cdot10^{2}$
Delta	$6.18\cdot10^{-5}$	$\mathbf{1.07\cdot10^{-6}}$	$6.70\cdot10^{-5}$	$1.19\cdot10^{-3}$	$5.19\cdot10^{-3}$	$1.18\cdot10^{-2}$	$1.26\cdot10^{3}$
Nothing	$\mathbf{3.64\cdot10^{-5}}$	$3.84\cdot10^{-5}$	$\mathbf{3.88\cdot10^{-5}}$	$\mathbf{-1.97\cdot10^{-4}}$	$\mathbf{2.47\cdot10^{-3}}$	$\mathbf{6.01\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.15 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$6.32 \cdot 10^{-3}$	$6.51 \cdot 10^{-3}$	$6.75 \cdot 10^{-3}$	$3.92 \cdot 10^{-3}$	$4.38 \cdot 10^{-2}$	$8.47 \cdot 10^{-2}$	$3.66 \cdot 10^{1}$
Reward 2	$1.06 \cdot 10^{-2}$	$8.12 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	$-5.38 \cdot 10^{-4}$	$2.31 \cdot 10^{-2}$	$1.28 \cdot 10^{-1}$	$6.59 \cdot 10^{1}$
Reward 3	$8.75 \cdot 10^{-3}$	$8.99 \cdot 10^{-3}$	$8.83 \cdot 10^{-3}$	$\mathbf{-5.40 \cdot 10^{-4}}$	$1.68 \cdot 10^{-2}$	$9.55 \cdot 10^{-2}$	$6.60 \cdot 10^{1}$
Reward 4	$1.10 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$	$1.18 \cdot 10^{-3}$	$1.71 \cdot 10^{-3}$	$1.98 \cdot 10^{-2}$	$2.95 \cdot 10^{-2}$	$4.32 \cdot 10^{1}$
Reward 5	$6.38 \cdot 10^{-3}$	$6.56 \cdot 10^{-3}$	$6.81 \cdot 10^{-3}$	$3.92 \cdot 10^{-3}$	$4.19 \cdot 10^{-2}$	$8.64 \cdot 10^{-2}$	$3.83 \cdot 10^{1}$
Reward 6	$1.05 \cdot 10^{-2}$	$8.12 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	$-5.38 \cdot 10^{-4}$	$2.31 \cdot 10^{-2}$	$1.26 \cdot 10^{-1}$	$6.59 \cdot 10^{1}$
Reward 7	$1.05 \cdot 10^{-2}$	$8.12 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	$-5.38 \cdot 10^{-4}$	$2.31 \cdot 10^{-2}$	$1.28 \cdot 10^{-1}$	$6.59 \cdot 10^{1}$
Delta	$\mathbf{3.31 \cdot 10^{-5}}$	$\mathbf{1.16 \cdot 10^{-6}}$	$\mathbf{3.55 \cdot 10^{-5}}$	$1.12 \cdot 10^{-3}$	$5.14 \cdot 10^{-3}$	$1.17 \cdot 10^{-2}$	$1.21 \cdot 10^{3}$
Nothing	$3.93 \cdot 10^{-5}$	$3.99 \cdot 10^{-5}$	$3.99 \cdot 10^{-5}$	$-1.68 \cdot 10^{-4}$	$\mathbf{2.75 \cdot 10^{-3}}$	$\mathbf{7.29 \cdot 10^{-3}}$	$\mathbf{0.00 \cdot 10^{0}}$

Deep model

Table 11.16 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.32{\cdot}10^{-4}$	$2.33{\cdot}10^{-4}$	$2.33{\cdot}10^{-4}$	$-2.16{\cdot}10^{-4}$	$4.07{\cdot}10^{-4}$	$1.53{\cdot}10^{-2}$	$3.90{\cdot}10^{1}$
Reward 2	$1.12{\cdot}10^{-4}$	$1.13{\cdot}10^{-4}$	$1.13{\cdot}10^{-4}$	$-1.79{\cdot}10^{-4}$	$7.59{\cdot}10^{-3}$	$1.51{\cdot}10^{-2}$	$5.61{\cdot}10^{4}$
Reward 3	$5.23{\cdot}10^{-4}$	$5.26{\cdot}10^{-4}$	$5.26{\cdot}10^{-4}$	$-3.26{\cdot}10^{-4}$	$7.88{\cdot}10^{-4}$	$2.30{\cdot}10^{-2}$	$1.36{\cdot}10^{1}$
Reward 4	$4.60{\cdot}10^{-4}$	$4.63{\cdot}10^{-4}$	$4.63{\cdot}10^{-4}$	$-3.06{\cdot}10^{-4}$	$1.11{\cdot}10^{-3}$	$2.16{\cdot}10^{-2}$	$7.82{\cdot}10^{3}$
Reward 5	$7.31{\cdot}10^{-7}$	$7.33{\cdot}10^{-7}$	$7.34{\cdot}10^{-7}$	$1.21{\cdot}10^{-6}$	$4.62{\cdot}10^{-4}$	$1.21{\cdot}10^{-3}$	$8.25{\cdot}10^{2}$
Reward 6	$1.09{\cdot}10^{-3}$	$1.10{\cdot}10^{-3}$	$1.10{\cdot}10^{-3}$	$\mathbf{-4.77{\cdot}10^{-4}}$	$1.48{\cdot}10^{-3}$	$3.32{\cdot}10^{-2}$	$6.69{\cdot}10^{0}$
Reward 7	$8.07{\cdot}10^{-4}$	$8.13{\cdot}10^{-4}$	$8.13{\cdot}10^{-4}$	$-3.44{\cdot}10^{-4}$	$1.38{\cdot}10^{-3}$	$2.27{\cdot}10^{-2}$	$1.35{\cdot}10^{5}$
Delta	$\mathbf{4.86{\cdot}10^{-9}}$	$\mathbf{4.90{\cdot}10^{-9}}$	$\mathbf{4.91{\cdot}10^{-9}}$	$2.08{\cdot}10^{-6}$	$\mathbf{1.89{\cdot}10^{-5}}$	$\mathbf{4.54{\cdot}10^{-5}}$	$1.23{\cdot}10^{3}$
Nothing	$9.82{\cdot}10^{-6}$	$9.88{\cdot}10^{-6}$	$9.88{\cdot}10^{-6}$	$-4.83{\cdot}10^{-5}$	$1.44{\cdot}10^{-3}$	$4.00{\cdot}10^{-3}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.17 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.51{\cdot}10^{-6}$	$1.58{\cdot}10^{-6}$	$1.59{\cdot}10^{-6}$	$-3.31{\cdot}10^{-5}$	$6.42{\cdot}10^{-4}$	$1.23{\cdot}10^{-3}$	$6.72{\cdot}10^{2}$
Reward 2	$3.36{\cdot}10^{-4}$	$3.55{\cdot}10^{-4}$	$3.56{\cdot}10^{-4}$	$\mathbf{-6.07{\cdot}10^{-4}}$	$3.63{\cdot}10^{-3}$	$1.19{\cdot}10^{-2}$	$5.65{\cdot}10^{3}$
Reward 3	$6.78{\cdot}10^{-6}$	$7.15{\cdot}10^{-6}$	$7.18{\cdot}10^{-6}$	$7.67{\cdot}10^{-5}$	$1.88{\cdot}10^{-3}$	$2.07{\cdot}10^{-3}$	$1.27{\cdot}10^{2}$
Reward 4	$2.18{\cdot}10^{-6}$	$2.27{\cdot}10^{-6}$	$2.28{\cdot}10^{-6}$	$-8.26{\cdot}10^{-6}$	$8.44{\cdot}10^{-4}$	$1.35{\cdot}10^{-3}$	$1.20{\cdot}10^{4}$
Reward 5	$4.52{\cdot}10^{-6}$	$4.77{\cdot}10^{-6}$	$4.79{\cdot}10^{-6}$	$-6.67{\cdot}10^{-5}$	$7.30{\cdot}10^{-4}$	$1.60{\cdot}10^{-3}$	$5.05{\cdot}10^{2}$
Reward 6	$3.29{\cdot}10^{-6}$	$3.46{\cdot}10^{-6}$	$3.47{\cdot}10^{-6}$	$5.53{\cdot}10^{-5}$	$1.52{\cdot}10^{-3}$	$1.69{\cdot}10^{-3}$	$2.02{\cdot}10^{2}$
Reward 7	$7.22{\cdot}10^{-7}$	$7.43{\cdot}10^{-7}$	$7.44{\cdot}10^{-7}$	$1.70{\cdot}10^{-5}$	$1.03{\cdot}10^{-3}$	$1.32{\cdot}10^{-3}$	$7.32{\cdot}10^{1}$
Delta	$\mathbf{7.61{\cdot}10^{-9}}$	$\mathbf{7.91{\cdot}10^{-9}}$	$\mathbf{7.93{\cdot}10^{-9}}$	$1.63{\cdot}10^{-6}$	$\mathbf{4.30{\cdot}10^{-5}}$	$\mathbf{5.56{\cdot}10^{-5}}$	$1.25{\cdot}10^{3}$
Nothing	$1.43{\cdot}10^{-4}$	$1.52{\cdot}10^{-4}$	$1.52{\cdot}10^{-4}$	$-3.91{\cdot}10^{-4}$	$2.01{\cdot}10^{-3}$	$7.44{\cdot}10^{-3}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.18 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$8.71 \cdot 10^{-7}$	$8.73 \cdot 10^{-7}$	$8.72 \cdot 10^{-7}$	$-4.82 \cdot 10^{-6}$	$7.07 \cdot 10^{-4}$	$1.47 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$
Reward 2	$3.51 \cdot 10^{-4}$	$3.51 \cdot 10^{-4}$	$3.52 \cdot 10^{-4}$	$\mathbf{-2.66 \cdot 10^{-4}}$	$5.57 \cdot 10^{-3}$	$2.45 \cdot 10^{-2}$	$3.81 \cdot 10^{4}$
Reward 3	$8.41 \cdot 10^{-7}$	$8.43 \cdot 10^{-7}$	$8.42 \cdot 10^{-7}$	$6.55 \cdot 10^{-6}$	$6.12 \cdot 10^{-4}$	$1.35 \cdot 10^{-3}$	$8.27 \cdot 10^{1}$
Reward 4	$6.57 \cdot 10^{-7}$	$6.58 \cdot 10^{-7}$	$6.58 \cdot 10^{-7}$	$2.25 \cdot 10^{-6}$	$5.32 \cdot 10^{-4}$	$1.21 \cdot 10^{-3}$	$4.18 \cdot 10^{2}$
Reward 5	$7.38 \cdot 10^{-7}$	$7.39 \cdot 10^{-7}$	$7.39 \cdot 10^{-7}$	$-1.28 \cdot 10^{-6}$	$7.12 \cdot 10^{-4}$	$1.36 \cdot 10^{-3}$	$2.54 \cdot 10^{-4}$
Reward 6	$8.66 \cdot 10^{-7}$	$8.67 \cdot 10^{-7}$	$8.67 \cdot 10^{-7}$	$-4.34 \cdot 10^{-6}$	$6.66 \cdot 10^{-4}$	$1.37 \cdot 10^{-3}$	$5.22 \cdot 10^{3}$
Reward 7	$7.37 \cdot 10^{-7}$	$7.39 \cdot 10^{-7}$	$7.38 \cdot 10^{-7}$	$1.26 \cdot 10^{-6}$	$6.78 \cdot 10^{-4}$	$1.35 \cdot 10^{-3}$	$1.13 \cdot 10^{2}$
Delta	$\mathbf{7.54 \cdot 10^{-10}}$	$\mathbf{7.55 \cdot 10^{-10}}$	$\mathbf{7.55 \cdot 10^{-10}}$	$1.51 \cdot 10^{-8}$	$\mathbf{8.78 \cdot 10^{-6}}$	$\mathbf{3.24 \cdot 10^{-5}}$	$1.24 \cdot 10^{3}$
Nothing	$1.09 \cdot 10^{-5}$	$1.09 \cdot 10^{-5}$	$1.09 \cdot 10^{-5}$	$-5.78 \cdot 10^{-5}$	$2.29 \cdot 10^{-3}$	$5.35 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.19 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$9.29 \cdot 10^{-7}$	$9.34 \cdot 10^{-7}$	$9.34 \cdot 10^{-7}$	$-1.04 \cdot 10^{-6}$	$6.23 \cdot 10^{-4}$	$1.27 \cdot 10^{-3}$	$1.31 \cdot 10^{3}$
Reward 2	$7.04 \cdot 10^{-4}$	$7.11 \cdot 10^{-4}$	$7.11 \cdot 10^{-4}$	$5.35 \cdot 10^{-5}$	$1.13 \cdot 10^{-2}$	$2.85 \cdot 10^{-2}$	$2.68 \cdot 10^{5}$
Reward 3	$1.18 \cdot 10^{-6}$	$1.18 \cdot 10^{-6}$	$1.18 \cdot 10^{-6}$	$7.62 \cdot 10^{-6}$	$6.01 \cdot 10^{-4}$	$1.22 \cdot 10^{-3}$	$1.97 \cdot 10^{2}$
Reward 4	$1.37 \cdot 10^{-6}$	$1.38 \cdot 10^{-6}$	$1.38 \cdot 10^{-6}$	$1.06 \cdot 10^{-5}$	$6.68 \cdot 10^{-4}$	$1.33 \cdot 10^{-3}$	$4.66 \cdot 10^{2}$
Reward 5	$9.60 \cdot 10^{-7}$	$9.66 \cdot 10^{-7}$	$9.66 \cdot 10^{-7}$	$-2.02 \cdot 10^{-6}$	$7.47 \cdot 10^{-4}$	$1.36 \cdot 10^{-3}$	$9.17 \cdot 10^{1}$
Reward 6	$1.44 \cdot 10^{-6}$	$1.45 \cdot 10^{-6}$	$1.45 \cdot 10^{-6}$	$9.86 \cdot 10^{-6}$	$7.07 \cdot 10^{-4}$	$1.39 \cdot 10^{-3}$	$5.56 \cdot 10^{3}$
Reward 7	$1.09 \cdot 10^{-6}$	$1.10 \cdot 10^{-6}$	$1.10 \cdot 10^{-6}$	$2.67 \cdot 10^{-6}$	$7.02 \cdot 10^{-4}$	$1.31 \cdot 10^{-3}$	$1.04 \cdot 10^{0}$
Delta	$\mathbf{3.95 \cdot 10^{-10}}$	$\mathbf{3.98 \cdot 10^{-10}}$	$\mathbf{4.01 \cdot 10^{-10}}$	$-1.59 \cdot 10^{-6}$	$\mathbf{3.80 \cdot 10^{-6}}$	$\mathbf{2.33 \cdot 10^{-5}}$	$1.22 \cdot 10^{3}$
Nothing	$9.40 \cdot 10^{-6}$	$9.47 \cdot 10^{-6}$	$9.49 \cdot 10^{-6}$	$\mathbf{-6.03 \cdot 10^{-5}}$	$2.61 \cdot 10^{-3}$	$4.85 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.20 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$4.84 \cdot 10^{-5}$	$4.85 \cdot 10^{-5}$	$4.85 \cdot 10^{-5}$	$1.26 \cdot 10^{-6}$	$1.22 \cdot 10^{-3}$	$8.75 \cdot 10^{-3}$	$3.92 \cdot 10^{6}$
Reward 2	$5.65 \cdot 10^{-5}$	$5.66 \cdot 10^{-5}$	$5.66 \cdot 10^{-5}$	$\mathbf{-1.07 \cdot 10^{-4}}$	$6.65 \cdot 10^{-3}$	$1.20 \cdot 10^{-2}$	$8.11 \cdot 10^{4}$
Reward 3	$7.20 \cdot 10^{-6}$	$7.22 \cdot 10^{-6}$	$7.22 \cdot 10^{-6}$	$-3.04 \cdot 10^{-5}$	$1.06 \cdot 10^{-3}$	$3.45 \cdot 10^{-3}$	$1.44 \cdot 10^{5}$
Reward 4	$2.52 \cdot 10^{-5}$	$2.53 \cdot 10^{-5}$	$2.53 \cdot 10^{-5}$	$-2.61 \cdot 10^{-5}$	$2.00 \cdot 10^{-3}$	$4.93 \cdot 10^{-3}$	$2.14 \cdot 10^{6}$
Reward 5	$3.14 \cdot 10^{-7}$	$3.14 \cdot 10^{-7}$	$3.14 \cdot 10^{-7}$	$-5.12 \cdot 10^{-6}$	$6.85 \cdot 10^{-4}$	$1.05 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$
Reward 6	$8.66 \cdot 10^{-5}$	$8.68 \cdot 10^{-5}$	$8.68 \cdot 10^{-5}$	$-5.67 \cdot 10^{-5}$	$1.55 \cdot 10^{-3}$	$1.22 \cdot 10^{-2}$	$1.38 \cdot 10^{4}$
Reward 7	$7.75 \cdot 10^{-7}$	$7.76 \cdot 10^{-7}$	$7.76 \cdot 10^{-7}$	$-6.22 \cdot 10^{-6}$	$5.61 \cdot 10^{-4}$	$1.43 \cdot 10^{-3}$	$2.92 \cdot 10^{2}$
Delta	$\mathbf{3.09 \cdot 10^{-10}}$	$\mathbf{3.09 \cdot 10^{-10}}$	$\mathbf{3.17 \cdot 10^{-10}}$	$-2.63 \cdot 10^{-6}$	$\mathbf{1.21 \cdot 10^{-9}}$	$\mathbf{2.08 \cdot 10^{-5}}$	$1.23 \cdot 10^{3}$
Nothing	$6.46 \cdot 10^{-6}$	$6.47 \cdot 10^{-6}$	$6.48 \cdot 10^{-6}$	$-4.37 \cdot 10^{-5}$	$2.78 \cdot 10^{-3}$	$4.84 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.21 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.10 \cdot 10^{-3}$	$1.13 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$	$1.13 \cdot 10^{-2}$	$3.64 \cdot 10^{-2}$	$\mathbf{0.00 \cdot 10^{0}}$
Reward 2	$1.10 \cdot 10^{-3}$	$1.13 \cdot 10^{-3}$	$1.14 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$	$1.13 \cdot 10^{-2}$	$3.64 \cdot 10^{-2}$	$\mathbf{0.00 \cdot 10^{0}}$
Reward 3	$3.67 \cdot 10^{-5}$	$3.73 \cdot 10^{-5}$	$6.17 \cdot 10^{-5}$	$3.90 \cdot 10^{-3}$	$1.63 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$1.37 \cdot 10^{5}$
Reward 4	$1.47 \cdot 10^{-3}$	$1.49 \cdot 10^{-3}$	$1.50 \cdot 10^{-3}$	$\mathbf{-9.99 \cdot 10^{-4}}$	$4.17 \cdot 10^{-3}$	$4.11 \cdot 10^{-2}$	$\mathbf{0.00 \cdot 10^{0}}$
Reward 5	$5.67 \cdot 10^{-7}$	$4.82 \cdot 10^{-7}$	$5.83 \cdot 10^{-7}$	$3.97 \cdot 10^{-5}$	$6.10 \cdot 10^{-4}$	$1.26 \cdot 10^{-3}$	$4.78 \cdot 10^{2}$
Reward 6	$6.43 \cdot 10^{-5}$	$6.62 \cdot 10^{-5}$	$6.68 \cdot 10^{-5}$	$2.73 \cdot 10^{-4}$	$3.94 \cdot 10^{-3}$	$8.10 \cdot 10^{-3}$	$2.71 \cdot 10^{2}$
Reward 7	$5.57 \cdot 10^{-6}$	$5.68 \cdot 10^{-6}$	$5.80 \cdot 10^{-6}$	$6.99 \cdot 10^{-5}$	$3.00 \cdot 10^{-3}$	$3.78 \cdot 10^{-3}$	$3.70 \cdot 10^{3}$
Delta	$\mathbf{8.49 \cdot 10^{-8}}$	$\mathbf{4.46 \cdot 10^{-9}}$	$\mathbf{9.18 \cdot 10^{-8}}$	$6.12 \cdot 10^{-5}$	$\mathbf{2.40 \cdot 10^{-4}}$	$\mathbf{6.12 \cdot 10^{-4}}$	$1.18 \cdot 10^{3}$
Nothing	$7.07 \cdot 10^{-6}$	$7.19 \cdot 10^{-6}$	$7.20 \cdot 10^{-6}$	$-7.70 \cdot 10^{-5}$	$1.48 \cdot 10^{-3}$	$3.95 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.22 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.62 \cdot 10^{-4}$	$2.64 \cdot 10^{-4}$	$2.64 \cdot 10^{-4}$	$2.93 \cdot 10^{-4}$	$3.02 \cdot 10^{-3}$	$1.78 \cdot 10^{-2}$	$8.83 \cdot 10^{1}$
Reward 2	$2.64 \cdot 10^{-3}$	$2.65 \cdot 10^{-3}$	$2.66 \cdot 10^{-3}$	$\mathbf{-7.28 \cdot 10^{-4}}$	$1.29 \cdot 10^{-2}$	$6.11 \cdot 10^{-2}$	$6.31 \cdot 10^{1}$
Reward 3	$9.08 \cdot 10^{-4}$	$9.13 \cdot 10^{-4}$	$9.13 \cdot 10^{-4}$	$-4.64 \cdot 10^{-4}$	$4.56 \cdot 10^{-3}$	$3.34 \cdot 10^{-2}$	$4.36 \cdot 10^{1}$
Reward 4	$7.47 \cdot 10^{-5}$	$7.51 \cdot 10^{-5}$	$7.52 \cdot 10^{-5}$	$1.58 \cdot 10^{-4}$	$2.61 \cdot 10^{-3}$	$9.27 \cdot 10^{-3}$	$1.41 \cdot 10^{3}$
Reward 5	$2.05 \cdot 10^{-6}$	$2.00 \cdot 10^{-6}$	$2.06 \cdot 10^{-6}$	$3.03 \cdot 10^{-5}$	$7.87 \cdot 10^{-4}$	$1.83 \cdot 10^{-3}$	$1.61 \cdot 10^{2}$
Reward 6	$4.74 \cdot 10^{-6}$	$4.75 \cdot 10^{-6}$	$4.77 \cdot 10^{-6}$	$2.29 \cdot 10^{-5}$	$1.03 \cdot 10^{-3}$	$2.75 \cdot 10^{-3}$	$2.49 \cdot 10^{3}$
Reward 7	$5.11 \cdot 10^{-6}$	$5.11 \cdot 10^{-6}$	$5.15 \cdot 10^{-6}$	$5.00 \cdot 10^{-5}$	$2.12 \cdot 10^{-3}$	$3.42 \cdot 10^{-3}$	$5.13 \cdot 10^{3}$
Delta	$\mathbf{5.69 \cdot 10^{-8}}$	$\mathbf{4.91 \cdot 10^{-9}}$	$\mathbf{6.05 \cdot 10^{-8}}$	$5.64 \cdot 10^{-5}$	$\mathbf{1.70 \cdot 10^{-4}}$	$\mathbf{5.62 \cdot 10^{-4}}$	$1.22 \cdot 10^{3}$
Nothing	$1.59 \cdot 10^{-5}$	$1.60 \cdot 10^{-5}$	$1.60 \cdot 10^{-5}$	$-7.70 \cdot 10^{-5}$	$1.98 \cdot 10^{-3}$	$5.59 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.23 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.89 \cdot 10^{-4}$	$1.89 \cdot 10^{-4}$	$1.89 \cdot 10^{-4}$	$-3.51 \cdot 10^{-5}$	$4.88 \cdot 10^{-3}$	$2.21 \cdot 10^{-2}$	$5.08 \cdot 10^{2}$
Reward 2	$1.27 \cdot 10^{-3}$	$1.28 \cdot 10^{-3}$	$1.29 \cdot 10^{-3}$	$5.47 \cdot 10^{-4}$	$2.34 \cdot 10^{-2}$	$4.46 \cdot 10^{-2}$	$1.35 \cdot 10^{2}$
Reward 3	$1.95 \cdot 10^{-4}$	$1.97 \cdot 10^{-4}$	$1.98 \cdot 10^{-4}$	$9.29 \cdot 10^{-5}$	$7.27 \cdot 10^{-3}$	$1.91 \cdot 10^{-2}$	$1.08 \cdot 10^{2}$
Reward 4	$1.26 \cdot 10^{-4}$	$1.28 \cdot 10^{-4}$	$1.29 \cdot 10^{-4}$	$2.01 \cdot 10^{-4}$	$5.94 \cdot 10^{-3}$	$8.36 \cdot 10^{-3}$	$1.05 \cdot 10^{3}$
Reward 5	$6.54 \cdot 10^{-7}$	$6.37 \cdot 10^{-7}$	$6.56 \cdot 10^{-7}$	$-1.64 \cdot 10^{-6}$	$1.06 \cdot 10^{-3}$	$1.60 \cdot 10^{-3}$	$6.38 \cdot 10^{2}$
Reward 6	$6.55 \cdot 10^{-5}$	$6.63 \cdot 10^{-5}$	$6.65 \cdot 10^{-5}$	$1.42 \cdot 10^{-4}$	$5.04 \cdot 10^{-3}$	$9.49 \cdot 10^{-3}$	$1.63 \cdot 10^{2}$
Reward 7	$8.10 \cdot 10^{-5}$	$8.20 \cdot 10^{-5}$	$8.21 \cdot 10^{-5}$	$\mathbf{-1.44 \cdot 10^{-4}}$	$2.72 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$	$9.90 \cdot 10^{1}$
Delta	$\mathbf{4.83 \cdot 10^{-8}}$	$\mathbf{3.53 \cdot 10^{-9}}$	$\mathbf{5.16 \cdot 10^{-8}}$	$5.24 \cdot 10^{-5}$	$\mathbf{1.88 \cdot 10^{-4}}$	$\mathbf{5.14 \cdot 10^{-4}}$	$1.26 \cdot 10^{3}$
Nothing	$1.15 \cdot 10^{-5}$	$1.16 \cdot 10^{-5}$	$1.16 \cdot 10^{-5}$	$-6.60 \cdot 10^{-5}$	$2.26 \cdot 10^{-3}$	$5.10 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.24 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.76 \cdot 10^{-4}$	$1.77 \cdot 10^{-4}$	$1.79 \cdot 10^{-4}$	$4.58 \cdot 10^{-4}$	$5.42 \cdot 10^{-3}$	$1.72 \cdot 10^{-2}$	$2.66 \cdot 10^{3}$
Reward 2	$6.59 \cdot 10^{-4}$	$6.60 \cdot 10^{-4}$	$6.63 \cdot 10^{-4}$	$4.63 \cdot 10^{-4}$	$1.41 \cdot 10^{-2}$	$3.71 \cdot 10^{-2}$	$\mathbf{0.00 \cdot 10^{0}}$
Reward 3	$9.34 \cdot 10^{-4}$	$9.37 \cdot 10^{-4}$	$9.39 \cdot 10^{-4}$	$\mathbf{-3.85 \cdot 10^{-4}}$	$7.93 \cdot 10^{-3}$	$3.91 \cdot 10^{-2}$	$3.88 \cdot 10^{1}$
Reward 4	$1.19 \cdot 10^{-4}$	$1.20 \cdot 10^{-4}$	$1.20 \cdot 10^{-4}$	$1.78 \cdot 10^{-4}$	$4.82 \cdot 10^{-3}$	$1.45 \cdot 10^{-2}$	$2.87 \cdot 10^{2}$
Reward 5	$6.20 \cdot 10^{-7}$	$5.80 \cdot 10^{-7}$	$6.24 \cdot 10^{-7}$	$3.18 \cdot 10^{-5}$	$7.37 \cdot 10^{-4}$	$1.40 \cdot 10^{-3}$	$6.83 \cdot 10^{2}$
Reward 6	$3.53 \cdot 10^{-5}$	$3.53 \cdot 10^{-5}$	$3.55 \cdot 10^{-5}$	$6.90 \cdot 10^{-5}$	$2.85 \cdot 10^{-3}$	$8.16 \cdot 10^{-3}$	$1.16 \cdot 10^{1}$
Reward 7	$3.89 \cdot 10^{-6}$	$3.88 \cdot 10^{-6}$	$3.90 \cdot 10^{-6}$	$9.81 \cdot 10^{-6}$	$2.56 \cdot 10^{-3}$	$3.84 \cdot 10^{-3}$	$2.26 \cdot 10^{3}$
Delta	$\mathbf{3.87 \cdot 10^{-8}}$	$\mathbf{3.85 \cdot 10^{-9}}$	$\mathbf{4.15 \cdot 10^{-8}}$	$5.09 \cdot 10^{-5}$	$\mathbf{1.62 \cdot 10^{-4}}$	$\mathbf{5.06 \cdot 10^{-4}}$	$1.24 \cdot 10^{3}$
Nothing	$8.54 \cdot 10^{-6}$	$8.58 \cdot 10^{-6}$	$8.59 \cdot 10^{-6}$	$-5.74 \cdot 10^{-5}$	$2.57 \cdot 10^{-3}$	$4.97 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.25 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.57 \cdot 10^{-6}$	$1.56 \cdot 10^{-6}$	$1.57 \cdot 10^{-6}$	$-2.43 \cdot 10^{-6}$	$2.23 \cdot 10^{-3}$	$2.95 \cdot 10^{-3}$	$9.34 \cdot 10^{3}$
Reward 2	$3.39 \cdot 10^{-4}$	$3.37 \cdot 10^{-4}$	$3.39 \cdot 10^{-4}$	$\mathbf{-4.56 \cdot 10^{-5}}$	$3.23 \cdot 10^{-3}$	$2.75 \cdot 10^{-2}$	$6.53 \cdot 10^{1}$
Reward 3	$3.39 \cdot 10^{-4}$	$3.37 \cdot 10^{-4}$	$3.39 \cdot 10^{-4}$	$\mathbf{-4.56 \cdot 10^{-5}}$	$3.23 \cdot 10^{-3}$	$2.75 \cdot 10^{-2}$	$6.53 \cdot 10^{1}$
Reward 4	$2.45 \cdot 10^{-4}$	$2.43 \cdot 10^{-4}$	$2.45 \cdot 10^{-4}$	$1.89 \cdot 10^{-4}$	$1.31 \cdot 10^{-2}$	$3.01 \cdot 10^{-2}$	$\mathbf{0.00 \cdot 10^{0}}$
Reward 5	$5.25 \cdot 10^{-7}$	$4.89 \cdot 10^{-7}$	$5.25 \cdot 10^{-7}$	$6.10 \cdot 10^{-6}$	$6.94 \cdot 10^{-4}$	$1.23 \cdot 10^{-3}$	$1.87 \cdot 10^{2}$
Reward 6	$3.39 \cdot 10^{-4}$	$3.37 \cdot 10^{-4}$	$3.39 \cdot 10^{-4}$	$\mathbf{-4.56 \cdot 10^{-5}}$	$3.23 \cdot 10^{-3}$	$2.75 \cdot 10^{-2}$	$6.53 \cdot 10^{1}$
Reward 7	$2.14 \cdot 10^{-4}$	$2.13 \cdot 10^{-4}$	$2.14 \cdot 10^{-4}$	$8.59 \cdot 10^{-5}$	$1.37 \cdot 10^{-3}$	$2.21 \cdot 10^{-2}$	$2.55 \cdot 10^{1}$
Delta	$\mathbf{3.64 \cdot 10^{-8}}$	$\mathbf{3.49 \cdot 10^{-9}}$	$\mathbf{3.91 \cdot 10^{-8}}$	$4.99 \cdot 10^{-5}$	$\mathbf{1.67 \cdot 10^{-4}}$	$\mathbf{5.22 \cdot 10^{-4}}$	$1.21 \cdot 10^{3}$
Nothing	$5.49 \cdot 10^{-6}$	$5.50 \cdot 10^{-6}$	$5.50 \cdot 10^{-6}$	$-3.17 \cdot 10^{-5}$	$2.96 \cdot 10^{-3}$	$4.73 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.26 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$4.26\cdot10^{-4}$	$3.76\cdot10^{-4}$	$4.27\cdot10^{-4}$	$2.28\cdot10^{-4}$	$5.35\cdot10^{-3}$	$2.81\cdot10^{-2}$	$5.12\cdot10^{1}$
Reward 2	$2.53\cdot10^{-3}$	$9.81\cdot10^{-4}$	$2.58\cdot10^{-3}$	$1.65\cdot10^{-3}$	$1.76\cdot10^{-2}$	$9.57\cdot10^{-2}$	**$0.00\cdot10^{0}$**
Reward 3	$1.61\cdot10^{-3}$	$1.61\cdot10^{-3}$	$1.61\cdot10^{-3}$	$2.60\cdot10^{-4}$	$1.60\cdot10^{-2}$	$5.29\cdot10^{-2}$	$1.27\cdot10^{1}$
Reward 4	$3.97\cdot10^{-4}$	$2.66\cdot10^{-4}$	$3.97\cdot10^{-4}$	$9.46\cdot10^{-5}$	$3.38\cdot10^{-3}$	$3.27\cdot10^{-2}$	$1.77\cdot10^{1}$
Reward 5	$3.27\cdot10^{-5}$	$5.23\cdot10^{-6}$	$3.66\cdot10^{-5}$	$1.05\cdot10^{-3}$	$6.81\cdot10^{-3}$	$1.30\cdot10^{-2}$	$1.27\cdot10^{3}$
Reward 6	$3.33\cdot10^{-4}$	$2.28\cdot10^{-4}$	$3.33\cdot10^{-4}$	$4.45\cdot10^{-5}$	$2.88\cdot10^{-3}$	$2.99\cdot10^{-2}$	$7.55\cdot10^{0}$
Reward 7	$8.46\cdot10^{-5}$	$8.39\cdot10^{-5}$	$8.55\cdot10^{-5}$	$2.34\cdot10^{-4}$	$7.25\cdot10^{-3}$	$1.19\cdot10^{-2}$	$7.50\cdot10^{1}$
Delta	$3.51\cdot10^{-5}$	**$1.46\cdot10^{-6}$**	$3.68\cdot10^{-5}$	$1.08\cdot10^{-3}$	$4.87\cdot10^{-3}$	$1.13\cdot10^{-2}$	$1.23\cdot10^{3}$
Nothing	**$1.40\cdot10^{-5}$**	$1.42\cdot10^{-5}$	**$1.42\cdot10^{-5}$**	**$-7.32\cdot10^{-5}$**	**$1.47\cdot10^{-3}$**	**$4.62\cdot10^{-3}$**	**$0.00\cdot10^{0}$**

Table 11.27 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.70\cdot10^{-3}$	$1.24\cdot10^{-3}$	$2.85\cdot10^{-3}$	$2.37\cdot10^{-3}$	$3.10\cdot10^{-2}$	$8.79\cdot10^{-2}$	$6.79\cdot10^{-4}$
Reward 2	$4.32\cdot10^{-3}$	$1.75\cdot10^{-3}$	$4.33\cdot10^{-3}$	$5.03\cdot10^{-4}$	$3.16\cdot10^{-2}$	$1.09\cdot10^{-1}$	**$0.00\cdot10^{0}$**
Reward 3	$2.59\cdot10^{-3}$	$2.38\cdot10^{-3}$	$2.59\cdot10^{-3}$	$3.46\cdot10^{-4}$	$2.12\cdot10^{-2}$	$7.57\cdot10^{-2}$	$9.94\cdot10^{0}$
Reward 4	$1.22\cdot10^{-5}$	$1.32\cdot10^{-6}$	$1.25\cdot10^{-5}$	$1.65\cdot10^{-4}$	$3.89\cdot10^{-3}$	$6.69\cdot10^{-3}$	$4.00\cdot10^{2}$
Reward 5	$3.83\cdot10^{-5}$	**$8.34\cdot10^{-7}$**	$3.97\cdot10^{-5}$	$3.92\cdot10^{-4}$	$3.80\cdot10^{-3}$	$1.19\cdot10^{-2}$	$2.20\cdot10^{2}$
Reward 6	**$1.04\cdot10^{-5}$**	$9.89\cdot10^{-6}$	**$1.06\cdot10^{-5}$**	$5.30\cdot10^{-5}$	$2.27\cdot10^{-3}$	$4.98\cdot10^{-3}$	$1.79\cdot10^{2}$
Reward 7	$3.26\cdot10^{-5}$	$1.33\cdot10^{-6}$	$3.39\cdot10^{-5}$	$3.46\cdot10^{-4}$	$3.56\cdot10^{-3}$	$1.08\cdot10^{-2}$	$1.66\cdot10^{2}$
Delta	$3.70\cdot10^{-5}$	$1.20\cdot10^{-6}$	$3.94\cdot10^{-5}$	$1.24\cdot10^{-3}$	$5.54\cdot10^{-3}$	$1.27\cdot10^{-2}$	$1.20\cdot10^{3}$
Nothing	$1.12\cdot10^{-5}$	$1.15\cdot10^{-5}$	$1.16\cdot10^{-5}$	**$-1.14\cdot10^{-4}$**	**$2.02\cdot10^{-3}$**	**$4.83\cdot10^{-3}$**	**$0.00\cdot10^{0}$**

Table 11.28 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.15\cdot10^{-4}$	$1.08\cdot10^{-4}$	$1.16\cdot10^{-4}$	$7.00\cdot10^{-4}$	$8.37\cdot10^{-3}$	$2.18\cdot10^{-2}$	$1.32\cdot10^{2}$
Reward 2	$9.14\cdot10^{-4}$	$3.42\cdot10^{-4}$	$9.17\cdot10^{-4}$	$7.62\cdot10^{-4}$	$7.60\cdot10^{-3}$	$7.01\cdot10^{-2}$	$1.01\cdot10^{-5}$
Reward 3	$1.22\cdot10^{-5}$	$1.07\cdot10^{-5}$	$1.26\cdot10^{-5}$	$3.29\cdot10^{-4}$	$3.72\cdot10^{-3}$	$7.57\cdot10^{-3}$	$3.20\cdot10^{3}$
Reward 4	$1.89\cdot10^{-4}$	$7.85\cdot10^{-5}$	$1.89\cdot10^{-4}$	$1.46\cdot10^{-4}$	$3.47\cdot10^{-3}$	$3.13\cdot10^{-2}$	$5.80\cdot10^{1}$
Reward 5	$1.54\cdot10^{-5}$	**$3.97\cdot10^{-7}$**	$1.56\cdot10^{-5}$	$4.44\cdot10^{-4}$	**$1.78\cdot10^{-3}$**	$1.02\cdot10^{-2}$	$4.03\cdot10^{2}$
Reward 6	$1.53\cdot10^{-5}$	$1.25\cdot10^{-5}$	$1.54\cdot10^{-5}$	$5.44\cdot10^{-6}$	$2.57\cdot10^{-3}$	$7.74\cdot10^{-3}$	$1.83\cdot10^{1}$
Reward 7	**$4.78\cdot10^{-6}$**	$4.78\cdot10^{-6}$	**$4.79\cdot10^{-6}$**	$-1.20\cdot10^{-5}$	$2.27\cdot10^{-3}$	**$4.16\cdot10^{-3}$**	$3.24\cdot10^{1}$
Delta	$1.46\cdot10^{-5}$	$1.06\cdot10^{-6}$	$1.57\cdot10^{-5}$	$1.03\cdot10^{-3}$	$2.71\cdot10^{-3}$	$1.00\cdot10^{-2}$	$1.24\cdot10^{3}$
Nothing	$4.96\cdot10^{-6}$	$4.97\cdot10^{-6}$	$4.97\cdot10^{-6}$	**$-4.00\cdot10^{-5}$**	$2.23\cdot10^{-3}$	$4.19\cdot10^{-3}$	**$0.00\cdot10^{0}$**

Table 11.29 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.42\cdot10^{-4}$	$1.08\cdot10^{-4}$	$1.42\cdot10^{-4}$	$1.87\cdot10^{-4}$	$5.95\cdot10^{-3}$	$2.00\cdot10^{-2}$	$8.39\cdot10^{1}$
Reward 2	$1.79\cdot10^{-4}$	$1.64\cdot10^{-4}$	$1.80\cdot10^{-4}$	$1.43\cdot10^{-4}$	$7.07\cdot10^{-3}$	$2.04\cdot10^{-2}$	$5.42\cdot10^{1}$
Reward 3	$1.30\cdot10^{-4}$	$8.38\cdot10^{-5}$	$1.31\cdot10^{-4}$	$5.13\cdot10^{-4}$	$6.51\cdot10^{-3}$	$2.05\cdot10^{-2}$	$2.29\cdot10^{2}$
Reward 4	$9.06\cdot10^{-4}$	$8.79\cdot10^{-4}$	$9.07\cdot10^{-4}$	$2.50\cdot10^{-4}$	$1.30\cdot10^{-2}$	$4.73\cdot10^{-2}$	$5.27\cdot10^{1}$
Reward 5	$1.81\cdot10^{-5}$	**$7.75\cdot10^{-7}$**	$1.84\cdot10^{-5}$	$3.42\cdot10^{-4}$	$2.66\cdot10^{-3}$	$9.71\cdot10^{-3}$	$2.71\cdot10^{2}$
Reward 6	**$7.40\cdot10^{-6}$**	$6.86\cdot10^{-6}$	**$7.41\cdot10^{-6}$**	$1.56\cdot10^{-5}$	$2.90\cdot10^{-3}$	**$5.05\cdot10^{-3}$**	$8.09\cdot10^{1}$
Reward 7	$1.36\cdot10^{-5}$	$5.49\cdot10^{-6}$	$1.36\cdot10^{-5}$	$7.46\cdot10^{-5}$	**$1.35\cdot10^{-3}$**	$7.45\cdot10^{-3}$	$2.18\cdot10^{3}$
Delta	$2.65\cdot10^{-5}$	$1.39\cdot10^{-6}$	$2.79\cdot10^{-5}$	$1.12\cdot10^{-3}$	$3.74\cdot10^{-3}$	$1.20\cdot10^{-2}$	$1.17\cdot10^{3}$
Nothing	$1.17\cdot10^{-5}$	$1.17\cdot10^{-5}$	$1.17\cdot10^{-5}$	**$-7.86\cdot10^{-5}$**	$2.64\cdot10^{-3}$	$5.54\cdot10^{-3}$	**$0.00\cdot10^{0}$**

Table 11.30 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.13 \cdot 10^{-3}$	$2.02 \cdot 10^{-3}$	$2.13 \cdot 10^{-3}$	$7.74 \cdot 10^{-5}$	$2.04 \cdot 10^{-2}$	$7.98 \cdot 10^{-2}$	$6.22 \cdot 10^{-1}$
Reward 2	$8.74 \cdot 10^{-5}$	$8.64 \cdot 10^{-5}$	$8.87 \cdot 10^{-5}$	$4.66 \cdot 10^{-4}$	$7.91 \cdot 10^{-3}$	$1.65 \cdot 10^{-2}$	$1.58 \cdot 10^{2}$
Reward 3	$2.13 \cdot 10^{-4}$	$2.10 \cdot 10^{-4}$	$3.45 \cdot 10^{-4}$	$3.54 \cdot 10^{-3}$	$1.72 \cdot 10^{-2}$	$2.06 \cdot 10^{-2}$	$2.63 \cdot 10^{4}$
Reward 4	$1.06 \cdot 10^{-5}$	$1.05 \cdot 10^{-5}$	$1.07 \cdot 10^{-5}$	$-6.07 \cdot 10^{-5}$	$2.57 \cdot 10^{-3}$	$\mathbf{5.74 \cdot 10^{-3}}$	$2.75 \cdot 10^{0}$
Reward 5	$1.73 \cdot 10^{-5}$	$1.32 \cdot 10^{-6}$	$1.75 \cdot 10^{-5}$	$2.56 \cdot 10^{-4}$	$2.42 \cdot 10^{-3}$	$9.59 \cdot 10^{-3}$	$4.27 \cdot 10^{2}$
Reward 6	$\mathbf{9.90 \cdot 10^{-6}}$	$9.89 \cdot 10^{-6}$	$\mathbf{9.93 \cdot 10^{-6}}$	$-6.06 \cdot 10^{-5}$	$2.86 \cdot 10^{-3}$	$5.76 \cdot 10^{-3}$	$7.75 \cdot 10^{0}$
Reward 7	$1.19 \cdot 10^{-5}$	$1.04 \cdot 10^{-5}$	$1.20 \cdot 10^{-5}$	$-3.25 \cdot 10^{-5}$	$\mathbf{1.80 \cdot 10^{-3}}$	$6.12 \cdot 10^{-3}$	$2.79 \cdot 10^{0}$
Delta	$2.29 \cdot 10^{-5}$	$\mathbf{1.30 \cdot 10^{-6}}$	$2.42 \cdot 10^{-5}$	$1.08 \cdot 10^{-3}$	$3.64 \cdot 10^{-3}$	$1.17 \cdot 10^{-2}$	$1.19 \cdot 10^{3}$
Nothing	$1.12 \cdot 10^{-5}$	$1.13 \cdot 10^{-5}$	$1.13 \cdot 10^{-5}$	$\mathbf{-7.84 \cdot 10^{-5}}$	$2.85 \cdot 10^{-3}$	$6.03 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Pre-trained model

Table 11.31 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0.05$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$\mathbf{3.06 \cdot 10^{-8}}$	$1.50 \cdot 10^{-8}$	$\mathbf{3.24 \cdot 10^{-8}}$	$4.08 \cdot 10^{-5}$	$\mathbf{1.54 \cdot 10^{-4}}$	$\mathbf{4.31 \cdot 10^{-4}}$	$9.14 \cdot 10^{2}$
Reward 5	$4.68 \cdot 10^{-8}$	$2.62 \cdot 10^{-8}$	$4.90 \cdot 10^{-8}$	$4.53 \cdot 10^{-5}$	$1.87 \cdot 10^{-4}$	$5.12 \cdot 10^{-4}$	$1.01 \cdot 10^{3}$
Reward 7	$1.13 \cdot 10^{-7}$	$9.22 \cdot 10^{-8}$	$1.15 \cdot 10^{-7}$	$4.37 \cdot 10^{-5}$	$1.82 \cdot 10^{-4}$	$6.35 \cdot 10^{-4}$	$7.48 \cdot 10^{2}$
Delta	$3.80 \cdot 10^{-8}$	$\mathbf{3.39 \cdot 10^{-9}}$	$4.12 \cdot 10^{-8}$	$5.55 \cdot 10^{-5}$	$1.66 \cdot 10^{-4}$	$5.24 \cdot 10^{-4}$	$1.28 \cdot 10^{3}$
Nothing	$8.35 \cdot 10^{-6}$	$8.36 \cdot 10^{-6}$	$8.36 \cdot 10^{-6}$	$\mathbf{-4.07 \cdot 10^{-5}}$	$1.56 \cdot 10^{-3}$	$4.19 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.32 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0.05$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$3.54 \cdot 10^{-8}$	$2.33 \cdot 10^{-8}$	$3.62 \cdot 10^{-8}$	$2.66 \cdot 10^{-5}$	$1.89 \cdot 10^{-4}$	$4.42 \cdot 10^{-4}$	$6.10 \cdot 10^{2}$
Reward 5	$3.17 \cdot 10^{-8}$	$9.72 \cdot 10^{-9}$	$3.39 \cdot 10^{-8}$	$4.41 \cdot 10^{-5}$	$1.80 \cdot 10^{-4}$	$4.64 \cdot 10^{-4}$	$9.85 \cdot 10^{2}$
Reward 7	$\mathbf{2.95 \cdot 10^{-8}}$	$1.31 \cdot 10^{-8}$	$\mathbf{3.09 \cdot 10^{-8}}$	$3.61 \cdot 10^{-5}$	$2.04 \cdot 10^{-4}$	$\mathbf{4.41 \cdot 10^{-4}}$	$8.71 \cdot 10^{2}$
Delta	$3.53 \cdot 10^{-8}$	$\mathbf{3.29 \cdot 10^{-9}}$	$3.82 \cdot 10^{-8}$	$5.16 \cdot 10^{-5}$	$\mathbf{1.46 \cdot 10^{-4}}$	$5.03 \cdot 10^{-4}$	$1.22 \cdot 10^{3}$
Nothing	$7.74 \cdot 10^{-6}$	$7.74 \cdot 10^{-6}$	$7.75 \cdot 10^{-6}$	$\mathbf{-4.60 \cdot 10^{-5}}$	$1.84 \cdot 10^{-3}$	$4.66 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.33 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0.05$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$6.00 \cdot 10^{-8}$	$4.38 \cdot 10^{-8}$	$6.16 \cdot 10^{-8}$	$3.43 \cdot 10^{-5}$	$1.86 \cdot 10^{-4}$	$4.76 \cdot 10^{-4}$	$6.66 \cdot 10^{2}$
Reward 5	$\mathbf{2.74 \cdot 10^{-8}}$	$4.67 \cdot 10^{-9}$	$\mathbf{2.94 \cdot 10^{-8}}$	$4.26 \cdot 10^{-5}$	$1.54 \cdot 10^{-4}$	$\mathbf{4.22 \cdot 10^{-4}}$	$1.00 \cdot 10^{3}$
Reward 7	$4.37 \cdot 10^{-8}$	$2.66 \cdot 10^{-8}$	$4.47 \cdot 10^{-8}$	$3.00 \cdot 10^{-5}$	$2.02 \cdot 10^{-4}$	$4.76 \cdot 10^{-4}$	$7.04 \cdot 10^{2}$
Delta	$4.25 \cdot 10^{-8}$	$\mathbf{2.25 \cdot 10^{-9}}$	$4.54 \cdot 10^{-8}$	$5.11 \cdot 10^{-5}$	$\mathbf{1.39 \cdot 10^{-4}}$	$5.16 \cdot 10^{-4}$	$1.22 \cdot 10^{3}$
Nothing	$5.29 \cdot 10^{-6}$	$5.30 \cdot 10^{-6}$	$5.30 \cdot 10^{-6}$	$\mathbf{-3.59 \cdot 10^{-5}}$	$2.19 \cdot 10^{-3}$	$3.98 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.34 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0.05$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$4.62 \cdot 10^{-8}$	$3.12 \cdot 10^{-8}$	$4.75 \cdot 10^{-8}$	$3.49 \cdot 10^{-5}$	$1.73 \cdot 10^{-4}$	$4.33 \cdot 10^{-4}$	$8.95 \cdot 10^{2}$
Reward 5	$\mathbf{3.03 \cdot 10^{-8}}$	$1.12 \cdot 10^{-8}$	$\mathbf{3.23 \cdot 10^{-8}}$	$4.18 \cdot 10^{-5}$	$1.93 \cdot 10^{-4}$	$\mathbf{4.25 \cdot 10^{-4}}$	$9.75 \cdot 10^{2}$
Reward 7	$2.41 \cdot 10^{-7}$	$2.31 \cdot 10^{-7}$	$2.42 \cdot 10^{-7}$	$2.30 \cdot 10^{-5}$	$2.22 \cdot 10^{-4}$	$7.00 \cdot 10^{-4}$	$7.44 \cdot 10^{2}$
Delta	$3.70 \cdot 10^{-8}$	$\mathbf{3.49 \cdot 10^{-9}}$	$3.98 \cdot 10^{-8}$	$5.01 \cdot 10^{-5}$	$\mathbf{1.67 \cdot 10^{-4}}$	$4.84 \cdot 10^{-4}$	$1.24 \cdot 10^{3}$
Nothing	$9.87 \cdot 10^{-6}$	$9.90 \cdot 10^{-6}$	$9.92 \cdot 10^{-6}$	$\mathbf{-6.90 \cdot 10^{-5}}$	$2.53 \cdot 10^{-3}$	$5.33 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.35 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0.05$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$\mathbf{4.21 \cdot 10^{-8}}$	$2.95 \cdot 10^{-8}$	$\mathbf{4.48 \cdot 10^{-8}}$	$2.56 \cdot 10^{-5}$	$\mathbf{1.96 \cdot 10^{-4}}$	$\mathbf{3.85 \cdot 10^{-4}}$	$6.07 \cdot 10^{2}$
Reward 5	$8.92 \cdot 10^{-8}$	$4.99 \cdot 10^{-8}$	$9.07 \cdot 10^{-8}$	$3.64 \cdot 10^{-5}$	$2.24 \cdot 10^{-4}$	$5.04 \cdot 10^{-4}$	$9.84 \cdot 10^{2}$
Reward 7	$7.21 \cdot 10^{-8}$	$3.82 \cdot 10^{-8}$	$7.35 \cdot 10^{-8}$	$3.32 \cdot 10^{-5}$	$2.32 \cdot 10^{-4}$	$4.64 \cdot 10^{-4}$	$9.08 \cdot 10^{2}$
Delta	$7.92 \cdot 10^{-8}$	$\mathbf{2.87 \cdot 10^{-9}}$	$8.36 \cdot 10^{-8}$	$5.16 \cdot 10^{-5}$	$2.40 \cdot 10^{-4}$	$5.41 \cdot 10^{-4}$	$1.29 \cdot 10^{3}$
Nothing	$1.15 \cdot 10^{-5}$	$1.17 \cdot 10^{-5}$	$1.17 \cdot 10^{-5}$	$\mathbf{-8.02 \cdot 10^{-5}}$	$2.77 \cdot 10^{-3}$	$5.31 \cdot 10^{-3}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.36 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 1$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$8.26\cdot10^{-6}$	$6.29\cdot10^{-6}$	$8.31\cdot10^{-6}$	$1.98\cdot10^{-4}$	$\mathbf{1.18\cdot10^{-3}}$	$5.44\cdot10^{-3}$	$1.13\cdot10^{4}$
Reward 5	$1.20\cdot10^{-5}$	$1.78\cdot10^{-6}$	$1.29\cdot10^{-5}$	$8.94\cdot10^{-4}$	$3.45\cdot10^{-3}$	$8.46\cdot10^{-3}$	$1.05\cdot10^{3}$
Reward 7	$1.42\cdot10^{-5}$	$2.48\cdot10^{-6}$	$1.45\cdot10^{-5}$	$3.74\cdot10^{-4}$	$1.97\cdot10^{-3}$	$8.97\cdot10^{-3}$	$1.97\cdot10^{2}$
Delta	$1.77\cdot10^{-5}$	$\mathbf{1.29\cdot10^{-6}}$	$1.90\cdot10^{-5}$	$1.04\cdot10^{-3}$	$3.77\cdot10^{-3}$	$1.02\cdot10^{-2}$	$1.25\cdot10^{3}$
Nothing	$\mathbf{7.01\cdot10^{-6}}$	$7.06\cdot10^{-6}$	$\mathbf{7.05\cdot10^{-6}}$	$\mathbf{-4.76\cdot10^{-5}}$	$1.48\cdot10^{-3}$	$\mathbf{3.87\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.37 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 1$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$8.30\cdot10^{-6}$	$3.39\cdot10^{-6}$	$8.37\cdot10^{-6}$	$1.57\cdot10^{-4}$	$2.02\cdot10^{-3}$	$6.54\cdot10^{-3}$	$1.63\cdot10^{2}$
Reward 5	$1.29\cdot10^{-5}$	$\mathbf{1.02\cdot10^{-6}}$	$1.37\cdot10^{-5}$	$8.84\cdot10^{-4}$	$2.63\cdot10^{-3}$	$9.57\cdot10^{-3}$	$1.04\cdot10^{3}$
Reward 7	$7.32\cdot10^{-6}$	$1.72\cdot10^{-6}$	$7.74\cdot10^{-6}$	$6.18\cdot10^{-4}$	$2.13\cdot10^{-3}$	$7.22\cdot10^{-3}$	$9.00\cdot10^{2}$
Delta	$1.65\cdot10^{-5}$	$1.11\cdot10^{-6}$	$1.76\cdot10^{-5}$	$1.04\cdot10^{-3}$	$2.96\cdot10^{-3}$	$1.08\cdot10^{-2}$	$1.22\cdot10^{3}$
Nothing	$\mathbf{5.13\cdot10^{-6}}$	$5.14\cdot10^{-6}$	$\mathbf{5.14\cdot10^{-6}}$	$\mathbf{-3.76\cdot10^{-5}}$	$\mathbf{1.91\cdot10^{-3}}$	$\mathbf{4.14\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.38 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 1$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$\mathbf{4.66\cdot10^{-6}}$	$1.80\cdot10^{-6}$	$\mathbf{4.70\cdot10^{-6}}$	$1.92\cdot10^{-4}$	$1.69\cdot10^{-3}$	$5.26\cdot10^{-3}$	$7.78\cdot10^{3}$
Reward 5	$1.38\cdot10^{-5}$	$2.05\cdot10^{-6}$	$1.46\cdot10^{-5}$	$8.93\cdot10^{-4}$	$2.58\cdot10^{-3}$	$9.68\cdot10^{-3}$	$8.96\cdot10^{2}$
Reward 7	$7.65\cdot10^{-6}$	$1.46\cdot10^{-6}$	$7.79\cdot10^{-6}$	$3.45\cdot10^{-4}$	$\mathbf{1.58\cdot10^{-3}}$	$7.08\cdot10^{-3}$	$4.35\cdot10^{2}$
Delta	$1.76\cdot10^{-5}$	$\mathbf{1.45\cdot10^{-6}}$	$1.89\cdot10^{-5}$	$1.07\cdot10^{-3}$	$2.79\cdot10^{-3}$	$1.10\cdot10^{-2}$	$1.18\cdot10^{3}$
Nothing	$7.20\cdot10^{-6}$	$7.20\cdot10^{-6}$	$7.21\cdot10^{-6}$	$\mathbf{-4.83\cdot10^{-5}}$	$2.28\cdot10^{-3}$	$\mathbf{5.02\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.39 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 1$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$\mathbf{5.88\cdot10^{-6}}$	$1.77\cdot10^{-6}$	$\mathbf{6.27\cdot10^{-6}}$	$5.85\cdot10^{-4}$	$2.08\cdot10^{-3}$	$6.23\cdot10^{-3}$	$6.46\cdot10^{2}$
Reward 5	$1.14\cdot10^{-5}$	$1.20\cdot10^{-6}$	$1.22\cdot10^{-5}$	$8.07\cdot10^{-4}$	$2.28\cdot10^{-3}$	$8.67\cdot10^{-3}$	$8.85\cdot10^{2}$
Reward 7	$9.01\cdot10^{-6}$	$1.79\cdot10^{-6}$	$9.09\cdot10^{-6}$	$2.58\cdot10^{-4}$	$\mathbf{1.09\cdot10^{-3}}$	$7.30\cdot10^{-3}$	$4.54\cdot10^{2}$
Delta	$1.79\cdot10^{-5}$	$\mathbf{8.94\cdot10^{-7}}$	$1.91\cdot10^{-5}$	$1.04\cdot10^{-3}$	$2.91\cdot10^{-3}$	$1.08\cdot10^{-2}$	$1.16\cdot10^{3}$
Nothing	$1.07\cdot10^{-5}$	$1.07\cdot10^{-5}$	$1.07\cdot10^{-5}$	$\mathbf{-5.70\cdot10^{-5}}$	$2.53\cdot10^{-3}$	$\mathbf{5.25\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

Table 11.40 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 1$ for the pre-trained models.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.54\cdot10^{-5}$	$1.16\cdot10^{-5}$	$1.55\cdot10^{-5}$	$2.50\cdot10^{-4}$	$\mathbf{1.86\cdot10^{-3}}$	$7.07\cdot10^{-3}$	$1.42\cdot10^{4}$
Reward 5	$1.22\cdot10^{-5}$	$1.56\cdot10^{-6}$	$1.30\cdot10^{-5}$	$8.23\cdot10^{-4}$	$3.25\cdot10^{-3}$	$8.54\cdot10^{-3}$	$9.47\cdot10^{2}$
Reward 7	$\mathbf{5.26\cdot10^{-6}}$	$2.16\cdot10^{-6}$	$\mathbf{5.42\cdot10^{-6}}$	$3.87\cdot10^{-4}$	$2.05\cdot10^{-3}$	$5.35\cdot10^{-3}$	$9.95\cdot10^{2}$
Delta	$1.98\cdot10^{-5}$	$\mathbf{1.11\cdot10^{-6}}$	$2.11\cdot10^{-5}$	$1.06\cdot10^{-3}$	$3.44\cdot10^{-3}$	$1.08\cdot10^{-2}$	$1.26\cdot10^{3}$
Nothing	$9.39\cdot10^{-6}$	$9.45\cdot10^{-6}$	$9.45\cdot10^{-6}$	$\mathbf{-5.53\cdot10^{-5}}$	$2.85\cdot10^{-3}$	$\mathbf{5.29\cdot10^{-3}}$	$\mathbf{0.00\cdot10^{0}}$

11.2 Environment 2 tables

Base reward 1

Table 11.41 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$5.97{\cdot}10^{-8}$	$5.97{\cdot}10^{-8}$	$5.97{\cdot}10^{-8}$	$2.04{\cdot}10^{-6}$	$1.77{\cdot}10^{-4}$	$4.20{\cdot}10^{-4}$	$4.74{\cdot}10^{2}$
Delta	$4.47{\cdot}10^{-8}$	$\mathbf{4.00{\cdot}10^{-8}}$	$4.48{\cdot}10^{-8}$	$\mathbf{6.41{\cdot}10^{-7}}$	$\mathbf{1.10{\cdot}10^{-4}}$	$3.84{\cdot}10^{-4}$	$6.02{\cdot}10^{3}$
Nothing	$\mathbf{4.25{\cdot}10^{-8}}$	$4.25{\cdot}10^{-8}$	$\mathbf{4.25{\cdot}10^{-8}}$	$1.35{\cdot}10^{-6}$	$2.16{\cdot}10^{-4}$	$\mathbf{3.45{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.42 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.96{\cdot}10^{-8}$	$2.96{\cdot}10^{-8}$	$2.96{\cdot}10^{-8}$	$3.53{\cdot}10^{-7}$	$3.13{\cdot}10^{-4}$	$3.86{\cdot}10^{-4}$	$9.83{\cdot}10^{2}$
Delta	$\mathbf{2.26{\cdot}10^{-8}}$	$\mathbf{1.78{\cdot}10^{-8}}$	$\mathbf{2.26{\cdot}10^{-8}}$	$\mathbf{2.56{\cdot}10^{-7}}$	$\mathbf{1.04{\cdot}10^{-4}}$	$\mathbf{3.23{\cdot}10^{-4}}$	$5.97{\cdot}10^{3}$
Nothing	$2.59{\cdot}10^{-8}$	$2.59{\cdot}10^{-8}$	$2.59{\cdot}10^{-8}$	$1.34{\cdot}10^{-6}$	$2.78{\cdot}10^{-4}$	$3.32{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.43 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$4.57 \cdot 10^{-8}$	$4.57 \cdot 10^{-8}$	$4.57 \cdot 10^{-8}$	$1.80 \cdot 10^{-6}$	$3.32 \cdot 10^{-4}$	$4.23 \cdot 10^{-4}$	$8.15 \cdot 10^{3}$
Delta	$\mathbf{3.31 \cdot 10^{-8}}$	$\mathbf{2.83 \cdot 10^{-8}}$	$\mathbf{3.31 \cdot 10^{-8}}$	$\mathbf{-4.26 \cdot 10^{-7}}$	$\mathbf{1.28 \cdot 10^{-4}}$	$\mathbf{3.40 \cdot 10^{-4}}$	$6.02 \cdot 10^{3}$
Nothing	$4.41 \cdot 10^{-8}$	$4.41 \cdot 10^{-8}$	$4.41 \cdot 10^{-8}$	$1.35 \cdot 10^{-6}$	$3.67 \cdot 10^{-4}$	$4.05 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.44 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$6.24 \cdot 10^{-8}$	$6.24 \cdot 10^{-8}$	$6.24 \cdot 10^{-8}$	$8.76 \cdot 10^{-7}$	$5.23 \cdot 10^{-4}$	$5.43 \cdot 10^{-4}$	$5.56 \cdot 10^{3}$
Delta	$\mathbf{3.50 \cdot 10^{-8}}$	$\mathbf{3.03 \cdot 10^{-8}}$	$\mathbf{3.51 \cdot 10^{-8}}$	$\mathbf{-2.41 \cdot 10^{-6}}$	$\mathbf{1.79 \cdot 10^{-4}}$	$\mathbf{3.99 \cdot 10^{-4}}$	$5.58 \cdot 10^{3}$
Nothing	$6.85 \cdot 10^{-8}$	$6.85 \cdot 10^{-8}$	$6.85 \cdot 10^{-8}$	$1.32 \cdot 10^{-6}$	$5.37 \cdot 10^{-4}$	$5.67 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.45 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.53 \cdot 10^{-7}$	$1.53 \cdot 10^{-7}$	$1.53 \cdot 10^{-7}$	$4.38 \cdot 10^{-7}$	$7.51 \cdot 10^{-4}$	$8.68 \cdot 10^{-4}$	$5.17 \cdot 10^{2}$
Delta	$\mathbf{5.26 \cdot 10^{-8}}$	$\mathbf{4.79 \cdot 10^{-8}}$	$\mathbf{5.27 \cdot 10^{-8}}$	$\mathbf{-3.61 \cdot 10^{-6}}$	$\mathbf{2.13 \cdot 10^{-4}}$	$\mathbf{4.42 \cdot 10^{-4}}$	$5.48 \cdot 10^{3}$
Nothing	$1.17 \cdot 10^{-7}$	$1.17 \cdot 10^{-7}$	$1.17 \cdot 10^{-7}$	$1.33 \cdot 10^{-6}$	$6.69 \cdot 10^{-4}$	$6.84 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.46 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	**$2.88 \cdot 10^{-8}$**	$2.88 \cdot 10^{-8}$	**$2.88 \cdot 10^{-8}$**	$1.09 \cdot 10^{-6}$	$1.89 \cdot 10^{-4}$	**$3.08 \cdot 10^{-4}$**	$3.04 \cdot 10^{3}$
Delta	$3.33 \cdot 10^{-8}$	**$2.80 \cdot 10^{-8}$**	$3.33 \cdot 10^{-8}$	**$6.14 \cdot 10^{-7}$**	**$9.37 \cdot 10^{-5}$**	$3.59 \cdot 10^{-4}$	$6.02 \cdot 10^{3}$
Nothing	$3.15 \cdot 10^{-8}$	$3.15 \cdot 10^{-8}$	$3.15 \cdot 10^{-8}$	$1.36 \cdot 10^{-6}$	$2.21 \cdot 10^{-4}$	$3.22 \cdot 10^{-4}$	**$0.00 \cdot 10^{0}$**

Table 11.47 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$3.02 \cdot 10^{-8}$	$3.02 \cdot 10^{-8}$	$3.02 \cdot 10^{-8}$	$1.00 \cdot 10^{-6}$	$3.11 \cdot 10^{-4}$	$3.76 \cdot 10^{-4}$	$2.59 \cdot 10^{3}$
Delta	**$2.79 \cdot 10^{-8}$**	**$2.26 \cdot 10^{-8}$**	**$2.79 \cdot 10^{-8}$**	**$3.98 \cdot 10^{-7}$**	**$1.06 \cdot 10^{-4}$**	**$3.43 \cdot 10^{-4}$**	$6.13 \cdot 10^{3}$
Nothing	$3.19 \cdot 10^{-8}$	$3.19 \cdot 10^{-8}$	$3.19 \cdot 10^{-8}$	$1.34 \cdot 10^{-6}$	$2.94 \cdot 10^{-4}$	$3.51 \cdot 10^{-4}$	**$0.00 \cdot 10^{0}$**

Table 11.48 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$7.58 \cdot 10^{-8}$	$7.58 \cdot 10^{-8}$	$7.58 \cdot 10^{-8}$	$1.22 \cdot 10^{-6}$	$4.87 \cdot 10^{-4}$	$5.79 \cdot 10^{-4}$	$6.97 \cdot 10^{2}$
Delta	**$4.45 \cdot 10^{-8}$**	**$3.93 \cdot 10^{-8}$**	**$4.45 \cdot 10^{-8}$**	**$-3.37 \cdot 10^{-7}$**	**$1.51 \cdot 10^{-4}$**	**$3.99 \cdot 10^{-4}$**	$5.87 \cdot 10^{3}$
Nothing	$5.86 \cdot 10^{-8}$	$5.86 \cdot 10^{-8}$	$5.86 \cdot 10^{-8}$	$1.36 \cdot 10^{-6}$	$3.97 \cdot 10^{-4}$	$4.62 \cdot 10^{-4}$	**$0.00 \cdot 10^{0}$**

Table 11.49 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$5.77 \cdot 10^{-8}$	$5.77 \cdot 10^{-8}$	$5.77 \cdot 10^{-8}$	**$-1.58 \cdot 10^{-6}$**	$3.90 \cdot 10^{-4}$	$4.78 \cdot 10^{-4}$	$6.89 \cdot 10^{3}$
Delta	**$3.94 \cdot 10^{-8}$**	**$3.41 \cdot 10^{-8}$**	**$3.94 \cdot 10^{-8}$**	$-1.13 \cdot 10^{-6}$	**$1.61 \cdot 10^{-4}$**	**$3.72 \cdot 10^{-4}$**	$5.90 \cdot 10^{3}$
Nothing	$6.41 \cdot 10^{-8}$	$6.41 \cdot 10^{-8}$	$6.41 \cdot 10^{-8}$	$1.34 \cdot 10^{-6}$	$4.61 \cdot 10^{-4}$	$4.81 \cdot 10^{-4}$	**$0.00 \cdot 10^{0}$**

Table 11.50 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.82 \cdot 10^{-7}$	$1.82 \cdot 10^{-7}$	$1.82 \cdot 10^{-7}$	$2.45 \cdot 10^{-6}$	$7.84 \cdot 10^{-4}$	$9.59 \cdot 10^{-4}$	$5.12 \cdot 10^{3}$
Delta	**$3.61 \cdot 10^{-8}$**	**$3.09 \cdot 10^{-8}$**	**$3.62 \cdot 10^{-8}$**	**$-2.23 \cdot 10^{-6}$**	**$1.91 \cdot 10^{-4}$**	**$3.76 \cdot 10^{-4}$**	$5.77 \cdot 10^{3}$
Nothing	$8.07 \cdot 10^{-8}$	$8.07 \cdot 10^{-8}$	$8.07 \cdot 10^{-8}$	$1.34 \cdot 10^{-6}$	$5.73 \cdot 10^{-4}$	$5.67 \cdot 10^{-4}$	**$0.00 \cdot 10^{0}$**

Table 11.51 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$4.45{\cdot}10^{-8}$	$4.45{\cdot}10^{-8}$	$4.45{\cdot}10^{-8}$	$1.36{\cdot}10^{-6}$	$2.28{\cdot}10^{-4}$	$4.08{\cdot}10^{-4}$	$2.44{\cdot}10^{2}$
Delta	$5.20{\cdot}10^{-8}$	$3.29{\cdot}10^{-8}$	$5.20{\cdot}10^{-8}$	$1.41{\cdot}10^{-6}$	$\mathbf{1.98{\cdot}10^{-4}}$	$5.37{\cdot}10^{-4}$	$6.01{\cdot}10^{3}$
Nothing	$\mathbf{3.24{\cdot}10^{-8}}$	$\mathbf{3.24{\cdot}10^{-8}}$	$\mathbf{3.24{\cdot}10^{-8}}$	$\mathbf{1.36{\cdot}10^{-6}}$	$2.13{\cdot}10^{-4}$	$\mathbf{3.30{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.52 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$\mathbf{3.38{\cdot}10^{-8}}$	$\mathbf{3.30{\cdot}10^{-8}}$	$\mathbf{3.38{\cdot}10^{-8}}$	$\mathbf{1.30{\cdot}10^{-6}}$	$3.09{\cdot}10^{-4}$	$3.91{\cdot}10^{-4}$	$1.29{\cdot}10^{3}$
Delta	$5.62{\cdot}10^{-8}$	$3.70{\cdot}10^{-8}$	$5.62{\cdot}10^{-8}$	$1.42{\cdot}10^{-6}$	$\mathbf{3.06{\cdot}10^{-4}}$	$5.66{\cdot}10^{-4}$	$5.89{\cdot}10^{3}$
Nothing	$3.67{\cdot}10^{-8}$	$3.67{\cdot}10^{-8}$	$3.67{\cdot}10^{-8}$	$1.35{\cdot}10^{-6}$	$3.11{\cdot}10^{-4}$	$\mathbf{3.81{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.53 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$4.24{\cdot}10^{-8}$	$\mathbf{4.14{\cdot}10^{-8}}$	$4.24{\cdot}10^{-8}$	$1.39{\cdot}10^{-6}$	$\mathbf{3.68{\cdot}10^{-4}}$	$4.35{\cdot}10^{-4}$	$2.16{\cdot}10^{3}$
Delta	$6.06{\cdot}10^{-8}$	$4.15{\cdot}10^{-8}$	$6.06{\cdot}10^{-8}$	$1.43{\cdot}10^{-6}$	$3.73{\cdot}10^{-4}$	$5.93{\cdot}10^{-4}$	$5.89{\cdot}10^{3}$
Nothing	$\mathbf{4.15{\cdot}10^{-8}}$	$4.15{\cdot}10^{-8}$	$\mathbf{4.15{\cdot}10^{-8}}$	$\mathbf{1.34{\cdot}10^{-6}}$	$3.77{\cdot}10^{-4}$	$\mathbf{4.23{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.54 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$\mathbf{7.28{\cdot}10^{-8}}$	$\mathbf{7.24{\cdot}10^{-8}}$	$\mathbf{7.28{\cdot}10^{-8}}$	$1.39{\cdot}10^{-6}$	$5.64{\cdot}10^{-4}$	$\mathbf{6.01{\cdot}10^{-4}}$	$1.56{\cdot}10^{3}$
Delta	$1.03{\cdot}10^{-7}$	$8.42{\cdot}10^{-8}$	$1.03{\cdot}10^{-7}$	$1.42{\cdot}10^{-6}$	$5.67{\cdot}10^{-4}$	$7.46{\cdot}10^{-4}$	$5.61{\cdot}10^{3}$
Nothing	$8.45{\cdot}10^{-8}$	$8.45{\cdot}10^{-8}$	$8.45{\cdot}10^{-8}$	$\mathbf{1.35{\cdot}10^{-6}}$	$\mathbf{5.60{\cdot}10^{-4}}$	$6.04{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.55 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$9.10{\cdot}10^{-8}$	$9.02{\cdot}10^{-8}$	$9.10{\cdot}10^{-8}$	$\mathbf{1.28{\cdot}10^{-6}}$	$\mathbf{6.09{\cdot}10^{-4}}$	$6.52{\cdot}10^{-4}$	$3.94{\cdot}10^{2}$
Delta	$9.99{\cdot}10^{-8}$	$\mathbf{8.09{\cdot}10^{-8}}$	$9.99{\cdot}10^{-8}$	$1.44{\cdot}10^{-6}$	$6.10{\cdot}10^{-4}$	$7.38{\cdot}10^{-4}$	$5.66{\cdot}10^{3}$
Nothing	$\mathbf{8.09{\cdot}10^{-8}}$	$8.09{\cdot}10^{-8}$	$\mathbf{8.09{\cdot}10^{-8}}$	$1.33{\cdot}10^{-6}$	$6.12{\cdot}10^{-4}$	$\mathbf{6.03{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Deep reward 1

Table 11.56 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.43{\cdot}10^{-7}$	$1.43{\cdot}10^{-7}$	$1.43{\cdot}10^{-7}$	$1.41{\cdot}10^{-6}$	$4.78{\cdot}10^{-4}$	$7.92{\cdot}10^{-4}$	$2.43{\cdot}10^{6}$
Delta	$3.69{\cdot}10^{-8}$	$\mathbf{3.22{\cdot}10^{-8}}$	$3.69{\cdot}10^{-8}$	$\mathbf{5.14{\cdot}10^{-7}}$	$\mathbf{1.04{\cdot}10^{-4}}$	$3.73{\cdot}10^{-4}$	$5.95{\cdot}10^{3}$
Nothing	$\mathbf{3.58{\cdot}10^{-8}}$	$3.58{\cdot}10^{-8}$	$\mathbf{3.58{\cdot}10^{-8}}$	$1.35{\cdot}10^{-6}$	$2.19{\cdot}10^{-4}$	$\mathbf{3.36{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.57 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$4.14{\cdot}10^{-7}$	$4.14{\cdot}10^{-7}$	$4.14{\cdot}10^{-7}$	$\mathbf{-2.05{\cdot}10^{-6}}$	$6.54{\cdot}10^{-4}$	$1.33{\cdot}10^{-3}$	$4.06{\cdot}10^{5}$
Delta	$\mathbf{2.65{\cdot}10^{-8}}$	$\mathbf{2.17{\cdot}10^{-8}}$	$\mathbf{2.65{\cdot}10^{-8}}$	$1.05{\cdot}10^{-7}$	$\mathbf{1.14{\cdot}10^{-4}}$	$\mathbf{3.39{\cdot}10^{-4}}$	$6.01{\cdot}10^{3}$
Nothing	$3.00{\cdot}10^{-8}$	$3.00{\cdot}10^{-8}$	$3.00{\cdot}10^{-8}$	$1.34{\cdot}10^{-6}$	$2.85{\cdot}10^{-4}$	$3.50{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.58 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$7.55 \cdot 10^{-8}$	$7.56 \cdot 10^{-8}$	$7.56 \cdot 10^{-8}$	$1.82 \cdot 10^{-6}$	$4.59 \cdot 10^{-4}$	$5.34 \cdot 10^{-4}$	$2.54 \cdot 10^{3}$
Delta	$\mathbf{5.03 \cdot 10^{-8}}$	$\mathbf{4.55 \cdot 10^{-8}}$	$\mathbf{5.03 \cdot 10^{-8}}$	$\mathbf{-1.03 \cdot 10^{-6}}$	$\mathbf{1.58 \cdot 10^{-4}}$	$\mathbf{4.05 \cdot 10^{-4}}$	$5.80 \cdot 10^{3}$
Nothing	$6.54 \cdot 10^{-8}$	$6.54 \cdot 10^{-8}$	$6.54 \cdot 10^{-8}$	$1.33 \cdot 10^{-6}$	$4.04 \cdot 10^{-4}$	$4.71 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.59 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.33 \cdot 10^{-7}$	$2.33 \cdot 10^{-7}$	$2.33 \cdot 10^{-7}$	$-2.77 \cdot 10^{-7}$	$6.27 \cdot 10^{-4}$	$1.06 \cdot 10^{-3}$	$3.78 \cdot 10^{6}$
Delta	$\mathbf{3.22 \cdot 10^{-8}}$	$\mathbf{2.74 \cdot 10^{-8}}$	$\mathbf{3.22 \cdot 10^{-8}}$	$\mathbf{-1.84 \cdot 10^{-6}}$	$\mathbf{1.76 \cdot 10^{-4}}$	$\mathbf{3.71 \cdot 10^{-4}}$	$5.62 \cdot 10^{3}$
Nothing	$6.31 \cdot 10^{-8}$	$6.32 \cdot 10^{-8}$	$6.31 \cdot 10^{-8}$	$1.33 \cdot 10^{-6}$	$5.41 \cdot 10^{-4}$	$5.52 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.60 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.42 \cdot 10^{-7}$	$2.42 \cdot 10^{-7}$	$2.42 \cdot 10^{-7}$	$4.16 \cdot 10^{-7}$	$6.85 \cdot 10^{-4}$	$1.09 \cdot 10^{-3}$	$5.96 \cdot 10^{6}$
Delta	$\mathbf{3.53 \cdot 10^{-8}}$	$\mathbf{3.05 \cdot 10^{-8}}$	$\mathbf{3.53 \cdot 10^{-8}}$	$\mathbf{-2.43 \cdot 10^{-6}}$	$\mathbf{1.85 \cdot 10^{-4}}$	$\mathbf{3.85 \cdot 10^{-4}}$	$5.65 \cdot 10^{3}$
Nothing	$8.09 \cdot 10^{-8}$	$8.09 \cdot 10^{-8}$	$8.09 \cdot 10^{-8}$	$1.33 \cdot 10^{-6}$	$5.95 \cdot 10^{-4}$	$5.99 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.61 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$2.48 \cdot 10^{-8}$	$2.48 \cdot 10^{-8}$	$2.48 \cdot 10^{-8}$	$\mathbf{-2.58 \cdot 10^{-6}}$	$2.19 \cdot 10^{-4}$	$3.20 \cdot 10^{-4}$	$4.22 \cdot 10^{3}$
Delta	$2.61 \cdot 10^{-8}$	$\mathbf{2.08 \cdot 10^{-8}}$	$2.61 \cdot 10^{-8}$	$9.30 \cdot 10^{-7}$	$\mathbf{8.71 \cdot 10^{-5}}$	$3.38 \cdot 10^{-4}$	$6.08 \cdot 10^{3}$
Nothing	$\mathbf{2.46 \cdot 10^{-8}}$	$2.46 \cdot 10^{-8}$	$\mathbf{2.46 \cdot 10^{-8}}$	$1.36 \cdot 10^{-6}$	$2.15 \cdot 10^{-4}$	$\mathbf{2.95 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.62 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$5.13 \cdot 10^{-8}$	$5.13 \cdot 10^{-8}$	$5.13 \cdot 10^{-8}$	$\mathbf{-1.34 \cdot 10^{-6}}$	$3.52 \cdot 10^{-4}$	$4.89 \cdot 10^{-4}$	$2.65 \cdot 10^{4}$
Delta	$\mathbf{3.79 \cdot 10^{-8}}$	$\mathbf{3.26 \cdot 10^{-8}}$	$\mathbf{3.79 \cdot 10^{-8}}$	$-1.44 \cdot 10^{-7}$	$\mathbf{1.19 \cdot 10^{-4}}$	$\mathbf{3.85 \cdot 10^{-4}}$	$5.90 \cdot 10^{3}$
Nothing	$4.19 \cdot 10^{-8}$	$4.19 \cdot 10^{-8}$	$4.19 \cdot 10^{-8}$	$1.35 \cdot 10^{-6}$	$2.98 \cdot 10^{-4}$	$3.88 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.63 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$4.15 \cdot 10^{-8}$	$4.14 \cdot 10^{-8}$	$4.15 \cdot 10^{-8}$	$\mathbf{-2.11 \cdot 10^{-6}}$	$2.91 \cdot 10^{-4}$	$4.71 \cdot 10^{-4}$	$1.87 \cdot 10^{3}$
Delta	$\mathbf{1.80 \cdot 10^{-8}}$	$\mathbf{1.27 \cdot 10^{-8}}$	$\mathbf{1.80 \cdot 10^{-8}}$	$-2.19 \cdot 10^{-7}$	$\mathbf{1.28 \cdot 10^{-4}}$	$\mathbf{3.05 \cdot 10^{-4}}$	$6.06 \cdot 10^{3}$
Nothing	$2.64 \cdot 10^{-8}$	$2.64 \cdot 10^{-8}$	$2.64 \cdot 10^{-8}$	$1.36 \cdot 10^{-6}$	$3.51 \cdot 10^{-4}$	$3.54 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.64 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$5.69 \cdot 10^{-8}$	$5.69 \cdot 10^{-8}$	$5.70 \cdot 10^{-8}$	$1.07 \cdot 10^{-6}$	$4.76 \cdot 10^{-4}$	$5.33 \cdot 10^{-4}$	$2.76 \cdot 10^{4}$
Delta	$\mathbf{3.55 \cdot 10^{-8}}$	$\mathbf{3.02 \cdot 10^{-8}}$	$\mathbf{3.55 \cdot 10^{-8}}$	$\mathbf{-1.38 \cdot 10^{-6}}$	$\mathbf{1.82 \cdot 10^{-4}}$	$\mathbf{3.89 \cdot 10^{-4}}$	$5.77 \cdot 10^{3}$
Nothing	$6.34 \cdot 10^{-8}$	$6.35 \cdot 10^{-8}$	$6.34 \cdot 10^{-8}$	$1.34 \cdot 10^{-6}$	$5.03 \cdot 10^{-4}$	$5.20 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.65 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$7.41 \cdot 10^{-8}$	$7.41 \cdot 10^{-8}$	$7.41 \cdot 10^{-8}$	$-7.50 \cdot 10^{-7}$	$5.01 \cdot 10^{-4}$	$5.72 \cdot 10^{-4}$	$7.20 \cdot 10^{3}$
Delta	$\mathbf{3.29 \cdot 10^{-8}}$	$\mathbf{2.76 \cdot 10^{-8}}$	$\mathbf{3.29 \cdot 10^{-8}}$	$\mathbf{-1.97 \cdot 10^{-6}}$	$\mathbf{1.88 \cdot 10^{-4}}$	$\mathbf{3.68 \cdot 10^{-4}}$	$5.76 \cdot 10^{3}$
Nothing	$7.49 \cdot 10^{-8}$	$7.49 \cdot 10^{-8}$	$7.49 \cdot 10^{-8}$	$1.34 \cdot 10^{-6}$	$5.68 \cdot 10^{-4}$	$5.50 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.66 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.89{\cdot}10^{-7}$	$1.88{\cdot}10^{-7}$	$1.89{\cdot}10^{-7}$	$\mathbf{1.11{\cdot}10^{-6}}$	$2.88{\cdot}10^{-4}$	$9.08{\cdot}10^{-4}$	$8.29{\cdot}10^{2}$
Delta	$7.45{\cdot}10^{-8}$	$5.54{\cdot}10^{-8}$	$7.45{\cdot}10^{-8}$	$1.41{\cdot}10^{-6}$	$\mathbf{1.93{\cdot}10^{-4}}$	$5.86{\cdot}10^{-4}$	$6.06{\cdot}10^{3}$
Nothing	$\mathbf{5.45{\cdot}10^{-8}}$	$\mathbf{5.45{\cdot}10^{-8}}$	$\mathbf{5.45{\cdot}10^{-8}}$	$1.35{\cdot}10^{-6}$	$2.12{\cdot}10^{-4}$	$\mathbf{3.75{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.67 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$9.76{\cdot}10^{-8}$	$9.44{\cdot}10^{-8}$	$9.76{\cdot}10^{-8}$	$1.68{\cdot}10^{-6}$	$4.28{\cdot}10^{-4}$	$7.02{\cdot}10^{-4}$	$1.20{\cdot}10^{4}$
Delta	$6.65{\cdot}10^{-8}$	$4.74{\cdot}10^{-8}$	$6.65{\cdot}10^{-8}$	$1.41{\cdot}10^{-6}$	$\mathbf{2.86{\cdot}10^{-4}}$	$5.69{\cdot}10^{-4}$	$5.95{\cdot}10^{3}$
Nothing	$\mathbf{4.71{\cdot}10^{-8}}$	$\mathbf{4.71{\cdot}10^{-8}}$	$\mathbf{4.71{\cdot}10^{-8}}$	$\mathbf{1.36{\cdot}10^{-6}}$	$2.95{\cdot}10^{-4}$	$\mathbf{3.75{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.68 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$5.41{\cdot}10^{-8}$	$5.32{\cdot}10^{-8}$	$5.41{\cdot}10^{-8}$	$\mathbf{1.22{\cdot}10^{-6}}$	$3.25{\cdot}10^{-4}$	$4.87{\cdot}10^{-4}$	$9.86{\cdot}10^{2}$
Delta	$4.93{\cdot}10^{-8}$	$\mathbf{3.01{\cdot}10^{-8}}$	$4.93{\cdot}10^{-8}$	$1.41{\cdot}10^{-6}$	$\mathbf{3.24{\cdot}10^{-4}}$	$5.29{\cdot}10^{-4}$	$6.16{\cdot}10^{3}$
Nothing	$\mathbf{3.04{\cdot}10^{-8}}$	$3.04{\cdot}10^{-8}$	$\mathbf{3.04{\cdot}10^{-8}}$	$1.36{\cdot}10^{-6}$	$3.26{\cdot}10^{-4}$	$\mathbf{3.38{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.69 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$8.48{\cdot}10^{-7}$	$8.45{\cdot}10^{-7}$	$8.48{\cdot}10^{-7}$	$\mathbf{1.29{\cdot}10^{-6}}$	$2.06{\cdot}10^{-3}$	$1.65{\cdot}10^{-3}$	$3.72{\cdot}10^{3}$
Delta	$6.62{\cdot}10^{-8}$	$\mathbf{4.71{\cdot}10^{-8}}$	$6.62{\cdot}10^{-8}$	$1.41{\cdot}10^{-6}$	$4.68{\cdot}10^{-4}$	$6.22{\cdot}10^{-4}$	$5.88{\cdot}10^{3}$
Nothing	$\mathbf{4.73{\cdot}10^{-8}}$	$4.74{\cdot}10^{-8}$	$\mathbf{4.73{\cdot}10^{-8}}$	$1.36{\cdot}10^{-6}$	$\mathbf{4.64{\cdot}10^{-4}}$	$\mathbf{4.58{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.70 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 1	$1.90{\cdot}10^{-7}$	$1.84{\cdot}10^{-7}$	$1.90{\cdot}10^{-7}$	$\mathbf{1.24{\cdot}10^{-6}}$	$8.09{\cdot}10^{-4}$	$9.90{\cdot}10^{-4}$	$3.66{\cdot}10^{3}$
Delta	$1.29{\cdot}10^{-7}$	$\mathbf{1.10{\cdot}10^{-7}}$	$1.29{\cdot}10^{-7}$	$1.46{\cdot}10^{-6}$	$7.16{\cdot}10^{-4}$	$8.34{\cdot}10^{-4}$	$5.39{\cdot}10^{3}$
Nothing	$\mathbf{1.10{\cdot}10^{-7}}$	$1.10{\cdot}10^{-7}$	$\mathbf{1.10{\cdot}10^{-7}}$	$1.31{\cdot}10^{-6}$	$\mathbf{7.00{\cdot}10^{-4}}$	$\mathbf{7.11{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Base reward 7

Table 11.71 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$2.41 \cdot 10^{-8}$	$2.41 \cdot 10^{-8}$	$2.41 \cdot 10^{-8}$	$1.64 \cdot 10^{-6}$	$2.10 \cdot 10^{-4}$	$3.05 \cdot 10^{-4}$	$4.06 \cdot 10^{3}$
Delta	$2.24 \cdot 10^{-8}$	$\mathbf{1.76 \cdot 10^{-8}}$	$2.24 \cdot 10^{-8}$	$\mathbf{9.42 \cdot 10^{-7}}$	$\mathbf{9.44 \cdot 10^{-5}}$	$3.14 \cdot 10^{-4}$	$6.10 \cdot 10^{3}$
Nothing	$\mathbf{2.09 \cdot 10^{-8}}$	$2.09 \cdot 10^{-8}$	$\mathbf{2.09 \cdot 10^{-8}}$	$1.36 \cdot 10^{-6}$	$2.15 \cdot 10^{-4}$	$\mathbf{2.78 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.72 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$3.85 \cdot 10^{-8}$	$3.85 \cdot 10^{-8}$	$3.85 \cdot 10^{-8}$	$1.09 \cdot 10^{-6}$	$3.17 \cdot 10^{-4}$	$4.09 \cdot 10^{-4}$	$2.20 \cdot 10^{2}$
Delta	$\mathbf{2.79 \cdot 10^{-8}}$	$\mathbf{2.31 \cdot 10^{-8}}$	$\mathbf{2.79 \cdot 10^{-8}}$	$\mathbf{-1.45 \cdot 10^{-7}}$	$\mathbf{1.33 \cdot 10^{-4}}$	$\mathbf{3.47 \cdot 10^{-4}}$	$5.95 \cdot 10^{3}$
Nothing	$3.33 \cdot 10^{-8}$	$3.33 \cdot 10^{-8}$	$3.33 \cdot 10^{-8}$	$1.35 \cdot 10^{-6}$	$3.17 \cdot 10^{-4}$	$3.74 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.73 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$3.57 \cdot 10^{-8}$	$3.57 \cdot 10^{-8}$	$3.57 \cdot 10^{-8}$	$3.04 \cdot 10^{-7}$	$2.84 \cdot 10^{-4}$	$3.96 \cdot 10^{-4}$	$4.26 \cdot 10^{3}$
Delta	$\mathbf{3.00 \cdot 10^{-8}}$	$\mathbf{2.53 \cdot 10^{-8}}$	$\mathbf{3.01 \cdot 10^{-8}}$	$\mathbf{-6.33 \cdot 10^{-7}}$	$\mathbf{1.36 \cdot 10^{-4}}$	$\mathbf{3.35 \cdot 10^{-4}}$	$5.95 \cdot 10^{3}$
Nothing	$3.95 \cdot 10^{-8}$	$3.95 \cdot 10^{-8}$	$3.95 \cdot 10^{-8}$	$1.35 \cdot 10^{-6}$	$3.52 \cdot 10^{-4}$	$3.87 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.74 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$1.20 \cdot 10^{-7}$	$1.20 \cdot 10^{-7}$	$1.21 \cdot 10^{-7}$	$6.68 \cdot 10^{-6}$	$6.85 \cdot 10^{-4}$	$6.81 \cdot 10^{-4}$	$3.56 \cdot 10^{3}$
Delta	$\mathbf{2.36 \cdot 10^{-8}}$	$\mathbf{1.88 \cdot 10^{-8}}$	$\mathbf{2.36 \cdot 10^{-8}}$	$\mathbf{-9.29 \cdot 10^{-7}}$	$\mathbf{1.28 \cdot 10^{-4}}$	$\mathbf{3.08 \cdot 10^{-4}}$	$5.93 \cdot 10^{3}$
Nothing	$4.05 \cdot 10^{-8}$	$4.05 \cdot 10^{-8}$	$4.05 \cdot 10^{-8}$	$1.36 \cdot 10^{-6}$	$4.03 \cdot 10^{-4}$	$4.02 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.75 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$1.36 \cdot 10^{-7}$	$1.36 \cdot 10^{-7}$	$1.36 \cdot 10^{-7}$	$3.13 \cdot 10^{-6}$	$7.85 \cdot 10^{-4}$	$8.30 \cdot 10^{-4}$	$7.00 \cdot 10^{2}$
Delta	$\mathbf{2.43 \cdot 10^{-8}}$	$\mathbf{1.95 \cdot 10^{-8}}$	$\mathbf{2.43 \cdot 10^{-8}}$	$\mathbf{-2.24 \cdot 10^{-6}}$	$\mathbf{1.64 \cdot 10^{-4}}$	$\mathbf{3.26 \cdot 10^{-4}}$	$5.75 \cdot 10^{3}$
Nothing	$6.41 \cdot 10^{-8}$	$6.41 \cdot 10^{-8}$	$6.41 \cdot 10^{-8}$	$1.34 \cdot 10^{-6}$	$5.50 \cdot 10^{-4}$	$5.30 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.76 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$3.35{\cdot}10^{-8}$	$3.35{\cdot}10^{-8}$	$3.35{\cdot}10^{-8}$	$1.28{\cdot}10^{-6}$	$2.01{\cdot}10^{-4}$	$3.47{\cdot}10^{-4}$	$3.03{\cdot}10^{2}$
Delta	$2.76{\cdot}10^{-8}$	$\mathbf{2.23{\cdot}10^{-8}}$	$2.76{\cdot}10^{-8}$	$\mathbf{5.99{\cdot}10^{-7}}$	$\mathbf{9.62{\cdot}10^{-5}}$	$3.45{\cdot}10^{-4}$	$5.95{\cdot}10^{3}$
Nothing	$\mathbf{2.66{\cdot}10^{-8}}$	$2.66{\cdot}10^{-8}$	$\mathbf{2.66{\cdot}10^{-8}}$	$1.36{\cdot}10^{-6}$	$2.32{\cdot}10^{-4}$	$\mathbf{3.09{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.77 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$3.00{\cdot}10^{-8}$	$3.00{\cdot}10^{-8}$	$3.00{\cdot}10^{-8}$	$6.60{\cdot}10^{-7}$	$3.16{\cdot}10^{-4}$	$3.72{\cdot}10^{-4}$	$2.66{\cdot}10^{3}$
Delta	$\mathbf{2.65{\cdot}10^{-8}}$	$\mathbf{2.12{\cdot}10^{-8}}$	$\mathbf{2.65{\cdot}10^{-8}}$	$\mathbf{3.10{\cdot}10^{-7}}$	$\mathbf{1.04{\cdot}10^{-4}}$	$3.20{\cdot}10^{-4}$	$6.08{\cdot}10^{3}$
Nothing	$2.85{\cdot}10^{-8}$	$2.85{\cdot}10^{-8}$	$2.85{\cdot}10^{-8}$	$1.36{\cdot}10^{-6}$	$2.69{\cdot}10^{-4}$	$\mathbf{3.13{\cdot}10^{-4}}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.78 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$1.52{\cdot}10^{-7}$	$1.52{\cdot}10^{-7}$	$1.52{\cdot}10^{-7}$	$1.24{\cdot}10^{-7}$	$4.82{\cdot}10^{-4}$	$5.43{\cdot}10^{-4}$	$6.95{\cdot}10^{3}$
Delta	$\mathbf{3.51{\cdot}10^{-8}}$	$\mathbf{2.98{\cdot}10^{-8}}$	$\mathbf{3.51{\cdot}10^{-8}}$	$\mathbf{-4.68{\cdot}10^{-7}}$	$\mathbf{1.43{\cdot}10^{-4}}$	$\mathbf{3.61{\cdot}10^{-4}}$	$5.94{\cdot}10^{3}$
Nothing	$4.75{\cdot}10^{-8}$	$4.75{\cdot}10^{-8}$	$4.75{\cdot}10^{-8}$	$1.35{\cdot}10^{-6}$	$3.85{\cdot}10^{-4}$	$4.19{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.79 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$4.81{\cdot}10^{-8}$	$4.81{\cdot}10^{-8}$	$4.81{\cdot}10^{-8}$	$1.35{\cdot}10^{-6}$	$4.08{\cdot}10^{-4}$	$4.59{\cdot}10^{-4}$	$3.26{\cdot}10^{2}$
Delta	$\mathbf{2.19{\cdot}10^{-8}}$	$\mathbf{1.67{\cdot}10^{-8}}$	$\mathbf{2.19{\cdot}10^{-8}}$	$\mathbf{-2.51{\cdot}10^{-7}}$	$\mathbf{1.28{\cdot}10^{-4}}$	$\mathbf{3.18{\cdot}10^{-4}}$	$6.04{\cdot}10^{3}$
Nothing	$3.92{\cdot}10^{-8}$	$3.92{\cdot}10^{-8}$	$3.92{\cdot}10^{-8}$	$1.35{\cdot}10^{-6}$	$4.18{\cdot}10^{-4}$	$4.13{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.80 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0.05$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$7.82{\cdot}10^{-8}$	$7.82{\cdot}10^{-8}$	$7.82{\cdot}10^{-8}$	$-1.20{\cdot}10^{-6}$	$4.32{\cdot}10^{-4}$	$5.08{\cdot}10^{-4}$	$1.22{\cdot}10^{4}$
Delta	$\mathbf{4.80{\cdot}10^{-8}}$	$\mathbf{4.28{\cdot}10^{-8}}$	$\mathbf{4.80{\cdot}10^{-8}}$	$\mathbf{-2.37{\cdot}10^{-6}}$	$\mathbf{1.91{\cdot}10^{-4}}$	$\mathbf{3.96{\cdot}10^{-4}}$	$5.77{\cdot}10^{3}$
Nothing	$9.15{\cdot}10^{-8}$	$9.15{\cdot}10^{-8}$	$9.15{\cdot}10^{-8}$	$1.32{\cdot}10^{-6}$	$5.49{\cdot}10^{-4}$	$5.65{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.81 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$5.48 \cdot 10^{-8}$	$5.46 \cdot 10^{-8}$	$5.48 \cdot 10^{-8}$	$\mathbf{1.33 \cdot 10^{-6}}$	$1.95 \cdot 10^{-4}$	$4.05 \cdot 10^{-4}$	$3.07 \cdot 10^{2}$
Delta	$6.49 \cdot 10^{-8}$	$4.58 \cdot 10^{-8}$	$6.49 \cdot 10^{-8}$	$1.42 \cdot 10^{-6}$	$\mathbf{1.89 \cdot 10^{-4}}$	$5.58 \cdot 10^{-4}$	$5.99 \cdot 10^{3}$
Nothing	$\mathbf{4.49 \cdot 10^{-8}}$	$\mathbf{4.50 \cdot 10^{-8}}$	$\mathbf{4.49 \cdot 10^{-8}}$	$1.34 \cdot 10^{-6}$	$2.03 \cdot 10^{-4}$	$\mathbf{3.48 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.82 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$4.81 \cdot 10^{-8}$	$4.71 \cdot 10^{-8}$	$4.81 \cdot 10^{-8}$	$\mathbf{1.30 \cdot 10^{-6}}$	$3.50 \cdot 10^{-4}$	$4.79 \cdot 10^{-4}$	$6.73 \cdot 10^{3}$
Delta	$6.52 \cdot 10^{-8}$	$\mathbf{4.61 \cdot 10^{-8}}$	$6.52 \cdot 10^{-8}$	$1.42 \cdot 10^{-6}$	$\mathbf{3.08 \cdot 10^{-4}}$	$5.89 \cdot 10^{-4}$	$5.89 \cdot 10^{3}$
Nothing	$\mathbf{4.61 \cdot 10^{-8}}$	$4.61 \cdot 10^{-8}$	$\mathbf{4.61 \cdot 10^{-8}}$	$1.34 \cdot 10^{-6}$	$3.16 \cdot 10^{-4}$	$\mathbf{4.07 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.83 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$5.56 \cdot 10^{-8}$	$5.53 \cdot 10^{-8}$	$5.56 \cdot 10^{-8}$	$\mathbf{1.20 \cdot 10^{-6}}$	$4.43 \cdot 10^{-4}$	$5.39 \cdot 10^{-4}$	$7.09 \cdot 10^{3}$
Delta	$6.42 \cdot 10^{-8}$	$\mathbf{4.51 \cdot 10^{-8}}$	$6.42 \cdot 10^{-8}$	$1.41 \cdot 10^{-6}$	$4.02 \cdot 10^{-4}$	$5.98 \cdot 10^{-4}$	$5.89 \cdot 10^{3}$
Nothing	$\mathbf{4.53 \cdot 10^{-8}}$	$4.53 \cdot 10^{-8}$	$\mathbf{4.53 \cdot 10^{-8}}$	$1.36 \cdot 10^{-6}$	$\mathbf{4.00 \cdot 10^{-4}}$	$\mathbf{4.28 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.84 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$7.53 \cdot 10^{-8}$	$6.45 \cdot 10^{-8}$	$7.53 \cdot 10^{-8}$	$1.48 \cdot 10^{-6}$	$\mathbf{4.50 \cdot 10^{-4}}$	$6.47 \cdot 10^{-4}$	$1.56 \cdot 10^{4}$
Delta	$7.83 \cdot 10^{-8}$	$\mathbf{5.92 \cdot 10^{-8}}$	$7.83 \cdot 10^{-8}$	$1.43 \cdot 10^{-6}$	$5.19 \cdot 10^{-4}$	$6.71 \cdot 10^{-4}$	$5.68 \cdot 10^{3}$
Nothing	$\mathbf{5.97 \cdot 10^{-8}}$	$5.97 \cdot 10^{-8}$	$\mathbf{5.97 \cdot 10^{-8}}$	$\mathbf{1.34 \cdot 10^{-6}}$	$5.05 \cdot 10^{-4}$	$\mathbf{5.21 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.85 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 1$.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$9.27 \cdot 10^{-8}$	$9.23 \cdot 10^{-8}$	$9.27 \cdot 10^{-8}$	$\mathbf{1.24 \cdot 10^{-6}}$	$6.11 \cdot 10^{-4}$	$6.73 \cdot 10^{-4}$	$2.79 \cdot 10^{2}$
Delta	$8.75 \cdot 10^{-8}$	$\mathbf{6.85 \cdot 10^{-8}}$	$8.75 \cdot 10^{-8}$	$1.42 \cdot 10^{-6}$	$5.74 \cdot 10^{-4}$	$7.06 \cdot 10^{-4}$	$5.72 \cdot 10^{3}$
Nothing	$\mathbf{6.87 \cdot 10^{-8}}$	$6.87 \cdot 10^{-8}$	$\mathbf{6.87 \cdot 10^{-8}}$	$1.34 \cdot 10^{-6}$	$\mathbf{5.74 \cdot 10^{-4}}$	$\mathbf{5.61 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Deep reward 7

Table 11.86 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$5.66 \cdot 10^{-7}$	$5.66 \cdot 10^{-7}$	$5.66 \cdot 10^{-7}$	$4.39 \cdot 10^{-6}$	$7.35 \cdot 10^{-4}$	$1.36 \cdot 10^{-3}$	$1.17 \cdot 10^{5}$
Delta	$2.68 \cdot 10^{-8}$	$\mathbf{2.21 \cdot 10^{-8}}$	$2.68 \cdot 10^{-8}$	$\mathbf{7.06 \cdot 10^{-7}}$	$\mathbf{1.07 \cdot 10^{-4}}$	$3.36 \cdot 10^{-4}$	$6.00 \cdot 10^{3}$
Nothing	$\mathbf{2.51 \cdot 10^{-8}}$	$2.51 \cdot 10^{-8}$	$\mathbf{2.51 \cdot 10^{-8}}$	$1.35 \cdot 10^{-6}$	$2.18 \cdot 10^{-4}$	$\mathbf{3.01 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.87 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$3.43 \cdot 10^{-7}$	$3.43 \cdot 10^{-7}$	$3.43 \cdot 10^{-7}$	$4.08 \cdot 10^{-6}$	$6.33 \cdot 10^{-4}$	$1.27 \cdot 10^{-3}$	$2.15 \cdot 10^{5}$
Delta	$\mathbf{3.02 \cdot 10^{-8}}$	$\mathbf{2.54 \cdot 10^{-8}}$	$\mathbf{3.02 \cdot 10^{-8}}$	$\mathbf{1.88 \cdot 10^{-7}}$	$\mathbf{1.22 \cdot 10^{-4}}$	$\mathbf{3.33 \cdot 10^{-4}}$	$5.94 \cdot 10^{3}$
Nothing	$3.32 \cdot 10^{-8}$	$3.32 \cdot 10^{-8}$	$3.32 \cdot 10^{-8}$	$1.36 \cdot 10^{-6}$	$2.84 \cdot 10^{-4}$	$3.41 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.88 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$9.98{\cdot}10^{-8}$	$9.98{\cdot}10^{-8}$	$9.98{\cdot}10^{-8}$	$2.77{\cdot}10^{-6}$	$5.13{\cdot}10^{-4}$	$7.01{\cdot}10^{-4}$	$3.54{\cdot}10^{5}$
Delta	$\mathbf{3.55{\cdot}10^{-8}}$	$\mathbf{3.07{\cdot}10^{-8}}$	$\mathbf{3.55{\cdot}10^{-8}}$	$\mathbf{-5.52{\cdot}10^{-7}}$	$\mathbf{1.37{\cdot}10^{-4}}$	$\mathbf{3.50{\cdot}10^{-4}}$	$5.96{\cdot}10^{3}$
Nothing	$4.70{\cdot}10^{-8}$	$4.70{\cdot}10^{-8}$	$4.70{\cdot}10^{-8}$	$1.35{\cdot}10^{-6}$	$3.76{\cdot}10^{-4}$	$4.17{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.89 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$2.51{\cdot}10^{-7}$	$2.51{\cdot}10^{-7}$	$2.51{\cdot}10^{-7}$	$-3.98{\cdot}10^{-7}$	$5.77{\cdot}10^{-4}$	$1.10{\cdot}10^{-3}$	$5.66{\cdot}10^{4}$
Delta	$\mathbf{3.86{\cdot}10^{-8}}$	$\mathbf{3.38{\cdot}10^{-8}}$	$\mathbf{3.86{\cdot}10^{-8}}$	$\mathbf{-1.19{\cdot}10^{-6}}$	$\mathbf{1.66{\cdot}10^{-4}}$	$\mathbf{3.93{\cdot}10^{-4}}$	$5.77{\cdot}10^{3}$
Nothing	$6.67{\cdot}10^{-8}$	$6.67{\cdot}10^{-8}$	$6.67{\cdot}10^{-8}$	$1.34{\cdot}10^{-6}$	$5.20{\cdot}10^{-4}$	$5.43{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.90 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$4.18{\cdot}10^{-7}$	$4.18{\cdot}10^{-7}$	$4.18{\cdot}10^{-7}$	$4.22{\cdot}10^{-6}$	$8.59{\cdot}10^{-4}$	$1.47{\cdot}10^{-3}$	$1.03{\cdot}10^{7}$
Delta	$\mathbf{2.43{\cdot}10^{-8}}$	$\mathbf{1.95{\cdot}10^{-8}}$	$\mathbf{2.43{\cdot}10^{-8}}$	$\mathbf{-2.23{\cdot}10^{-6}}$	$\mathbf{1.71{\cdot}10^{-4}}$	$\mathbf{3.29{\cdot}10^{-4}}$	$5.76{\cdot}10^{3}$
Nothing	$6.48{\cdot}10^{-8}$	$6.48{\cdot}10^{-8}$	$6.48{\cdot}10^{-8}$	$1.34{\cdot}10^{-6}$	$5.76{\cdot}10^{-4}$	$5.45{\cdot}10^{-4}$	$\mathbf{0.00{\cdot}10^{0}}$

Table 11.91 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$6.97 \cdot 10^{-8}$	$6.97 \cdot 10^{-8}$	$6.97 \cdot 10^{-8}$	$9.28 \cdot 10^{-7}$	$3.03 \cdot 10^{-4}$	$5.68 \cdot 10^{-4}$	$3.46 \cdot 10^{3}$
Delta	$3.04 \cdot 10^{-8}$	$\mathbf{2.51 \cdot 10^{-8}}$	$3.04 \cdot 10^{-8}$	$\mathbf{7.62 \cdot 10^{-7}}$	$\mathbf{8.16 \cdot 10^{-5}}$	$3.30 \cdot 10^{-4}$	$6.08 \cdot 10^{3}$
Nothing	$\mathbf{2.86 \cdot 10^{-8}}$	$2.86 \cdot 10^{-8}$	$\mathbf{2.86 \cdot 10^{-8}}$	$1.36 \cdot 10^{-6}$	$2.12 \cdot 10^{-4}$	$\mathbf{2.82 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.92 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$5.81 \cdot 10^{-8}$	$5.81 \cdot 10^{-8}$	$5.81 \cdot 10^{-8}$	$\mathbf{-1.74 \cdot 10^{-7}}$	$3.30 \cdot 10^{-4}$	$5.20 \cdot 10^{-4}$	$1.87 \cdot 10^{4}$
Delta	$\mathbf{2.59 \cdot 10^{-8}}$	$\mathbf{2.07 \cdot 10^{-8}}$	$\mathbf{2.59 \cdot 10^{-8}}$	$-5.37 \cdot 10^{-8}$	$\mathbf{1.11 \cdot 10^{-4}}$	$\mathbf{3.28 \cdot 10^{-4}}$	$6.01 \cdot 10^{3}$
Nothing	$2.92 \cdot 10^{-8}$	$2.92 \cdot 10^{-8}$	$2.92 \cdot 10^{-8}$	$1.36 \cdot 10^{-6}$	$2.89 \cdot 10^{-4}$	$3.34 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.93 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$5.84 \cdot 10^{-8}$	$5.84 \cdot 10^{-8}$	$5.84 \cdot 10^{-8}$	$9.08 \cdot 10^{-7}$	$3.96 \cdot 10^{-4}$	$5.38 \cdot 10^{-4}$	$9.18 \cdot 10^{3}$
Delta	$\mathbf{3.51 \cdot 10^{-8}}$	$\mathbf{2.99 \cdot 10^{-8}}$	$\mathbf{3.51 \cdot 10^{-8}}$	$\mathbf{-4.19 \cdot 10^{-7}}$	$\mathbf{1.39 \cdot 10^{-4}}$	$\mathbf{3.68 \cdot 10^{-4}}$	$6.06 \cdot 10^{3}$
Nothing	$4.59 \cdot 10^{-8}$	$4.59 \cdot 10^{-8}$	$4.59 \cdot 10^{-8}$	$1.34 \cdot 10^{-6}$	$3.76 \cdot 10^{-4}$	$4.18 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.94 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$5.58 \cdot 10^{-8}$	$5.58 \cdot 10^{-8}$	$5.58 \cdot 10^{-8}$	$1.08 \cdot 10^{-6}$	$4.06 \cdot 10^{-4}$	$5.17 \cdot 10^{-4}$	$1.96 \cdot 10^{4}$
Delta	$\mathbf{2.85 \cdot 10^{-8}}$	$\mathbf{2.32 \cdot 10^{-8}}$	$\mathbf{2.85 \cdot 10^{-8}}$	$\mathbf{-1.48 \cdot 10^{-6}}$	$\mathbf{1.71 \cdot 10^{-4}}$	$\mathbf{3.64 \cdot 10^{-4}}$	$5.83 \cdot 10^{3}$
Nothing	$5.62 \cdot 10^{-8}$	$5.62 \cdot 10^{-8}$	$5.62 \cdot 10^{-8}$	$1.34 \cdot 10^{-6}$	$5.00 \cdot 10^{-4}$	$5.06 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.95 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 0.05$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$1.43 \cdot 10^{-7}$	$1.43 \cdot 10^{-7}$	$1.43 \cdot 10^{-7}$	$2.20 \cdot 10^{-6}$	$7.29 \cdot 10^{-4}$	$8.35 \cdot 10^{-4}$	$1.06 \cdot 10^{3}$
Delta	$\mathbf{3.43 \cdot 10^{-8}}$	$\mathbf{2.91 \cdot 10^{-8}}$	$\mathbf{3.44 \cdot 10^{-8}}$	$\mathbf{-2.25 \cdot 10^{-6}}$	$\mathbf{2.02 \cdot 10^{-4}}$	$\mathbf{3.72 \cdot 10^{-4}}$	$5.62 \cdot 10^{3}$
Nothing	$7.98 \cdot 10^{-8}$	$7.98 \cdot 10^{-8}$	$7.98 \cdot 10^{-8}$	$1.33 \cdot 10^{-6}$	$6.06 \cdot 10^{-4}$	$5.81 \cdot 10^{-4}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.96 Average performance metrics over 100 realizations with increment correlation $\rho = -1$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$2.09 \cdot 10^{-7}$	$2.09 \cdot 10^{-7}$	$2.09 \cdot 10^{-7}$	$\mathbf{1.32 \cdot 10^{-6}}$	$5.56 \cdot 10^{-4}$	$9.63 \cdot 10^{-4}$	$1.89 \cdot 10^{4}$
Delta	$5.13 \cdot 10^{-8}$	$\mathbf{3.22 \cdot 10^{-8}}$	$5.13 \cdot 10^{-8}$	$1.41 \cdot 10^{-6}$	$\mathbf{2.04 \cdot 10^{-4}}$	$5.32 \cdot 10^{-4}$	$6.12 \cdot 10^{3}$
Nothing	$\mathbf{3.28 \cdot 10^{-8}}$	$3.28 \cdot 10^{-8}$	$\mathbf{3.28 \cdot 10^{-8}}$	$1.36 \cdot 10^{-6}$	$2.20 \cdot 10^{-4}$	$\mathbf{3.24 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.97 Average performance metrics over 100 realizations with increment correlation $\rho = -0.5$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$5.59 \cdot 10^{-8}$	$5.50 \cdot 10^{-8}$	$5.59 \cdot 10^{-8}$	$\mathbf{1.33 \cdot 10^{-6}}$	$3.92 \cdot 10^{-4}$	$5.22 \cdot 10^{-4}$	$1.68 \cdot 10^{4}$
Delta	$5.02 \cdot 10^{-8}$	$\mathbf{3.10 \cdot 10^{-8}}$	$5.02 \cdot 10^{-8}$	$1.40 \cdot 10^{-6}$	$\mathbf{2.75 \cdot 10^{-4}}$	$5.30 \cdot 10^{-4}$	$6.06 \cdot 10^{3}$
Nothing	$\mathbf{3.13 \cdot 10^{-8}}$	$3.13 \cdot 10^{-8}$	$\mathbf{3.13 \cdot 10^{-8}}$	$1.36 \cdot 10^{-6}$	$2.82 \cdot 10^{-4}$	$\mathbf{3.34 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.98 Average performance metrics over 100 realizations with increment correlation $\rho = 0$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$7.55 \cdot 10^{-8}$	$7.49 \cdot 10^{-8}$	$7.55 \cdot 10^{-8}$	$\mathbf{1.22 \cdot 10^{-6}}$	$4.06 \cdot 10^{-4}$	$6.05 \cdot 10^{-4}$	$8.34 \cdot 10^{3}$
Delta	$5.61 \cdot 10^{-8}$	$\mathbf{3.70 \cdot 10^{-8}}$	$5.61 \cdot 10^{-8}$	$1.42 \cdot 10^{-6}$	$\mathbf{3.61 \cdot 10^{-4}}$	$5.67 \cdot 10^{-4}$	$6.03 \cdot 10^{3}$
Nothing	$\mathbf{3.71 \cdot 10^{-8}}$	$3.71 \cdot 10^{-8}$	$\mathbf{3.71 \cdot 10^{-8}}$	$1.35 \cdot 10^{-6}$	$3.64 \cdot 10^{-4}$	$\mathbf{3.86 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.99 Average performance metrics over 100 realizations with increment correlation $\rho = 0.5$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$1.11 \cdot 10^{-7}$	$1.07 \cdot 10^{-7}$	$1.11 \cdot 10^{-7}$	$\mathbf{1.18 \cdot 10^{-6}}$	$5.81 \cdot 10^{-4}$	$7.60 \cdot 10^{-4}$	$6.01 \cdot 10^{3}$
Delta	$9.73 \cdot 10^{-8}$	$7.82 \cdot 10^{-8}$	$9.73 \cdot 10^{-8}$	$1.44 \cdot 10^{-6}$	$5.09 \cdot 10^{-4}$	$6.96 \cdot 10^{-4}$	$5.75 \cdot 10^{3}$
Nothing	$\mathbf{7.72 \cdot 10^{-8}}$	$\mathbf{7.72 \cdot 10^{-8}}$	$\mathbf{7.72 \cdot 10^{-8}}$	$1.33 \cdot 10^{-6}$	$\mathbf{5.04 \cdot 10^{-4}}$	$\mathbf{5.38 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

Table 11.100 Average performance metrics over 100 realizations with increment correlation $\rho = 1$ and trading cost $c = 1$ and deep network architecture.* excluding first point.

	Variance of Losses	Variance of Losses*	Mean Square Losses	Mean Losses	Value at Risk	Expected Shortfall	Turnover
Reward 7	$1.44 \cdot 10^{-7}$	$1.44 \cdot 10^{-7}$	$1.44 \cdot 10^{-7}$	$1.56 \cdot 10^{-6}$	$7.38 \cdot 10^{-4}$	$8.76 \cdot 10^{-4}$	$2.05 \cdot 10^{4}$
Delta	$9.32 \cdot 10^{-8}$	$\mathbf{7.41 \cdot 10^{-8}}$	$9.32 \cdot 10^{-8}$	$1.44 \cdot 10^{-6}$	$6.04 \cdot 10^{-4}$	$7.25 \cdot 10^{-4}$	$5.60 \cdot 10^{3}$
Nothing	$\mathbf{7.43 \cdot 10^{-8}}$	$7.43 \cdot 10^{-8}$	$\mathbf{7.43 \cdot 10^{-8}}$	$\mathbf{1.32 \cdot 10^{-6}}$	$\mathbf{5.95 \cdot 10^{-4}}$	$\mathbf{5.82 \cdot 10^{-4}}$	$\mathbf{0.00 \cdot 10^{0}}$

11.3 Environment 1 figures

Figure 11.1 Example realization of environment 1 with no correlation.

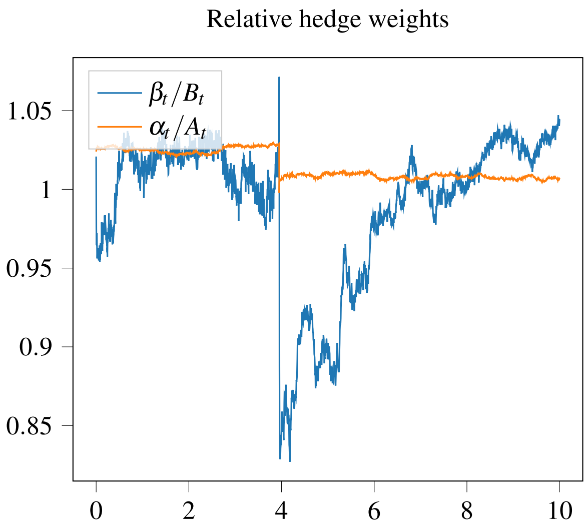

Figure 11.2 Example trading strategy on environment 1 by an RL agent trained on reward 1 with no correlation and no trading cost.

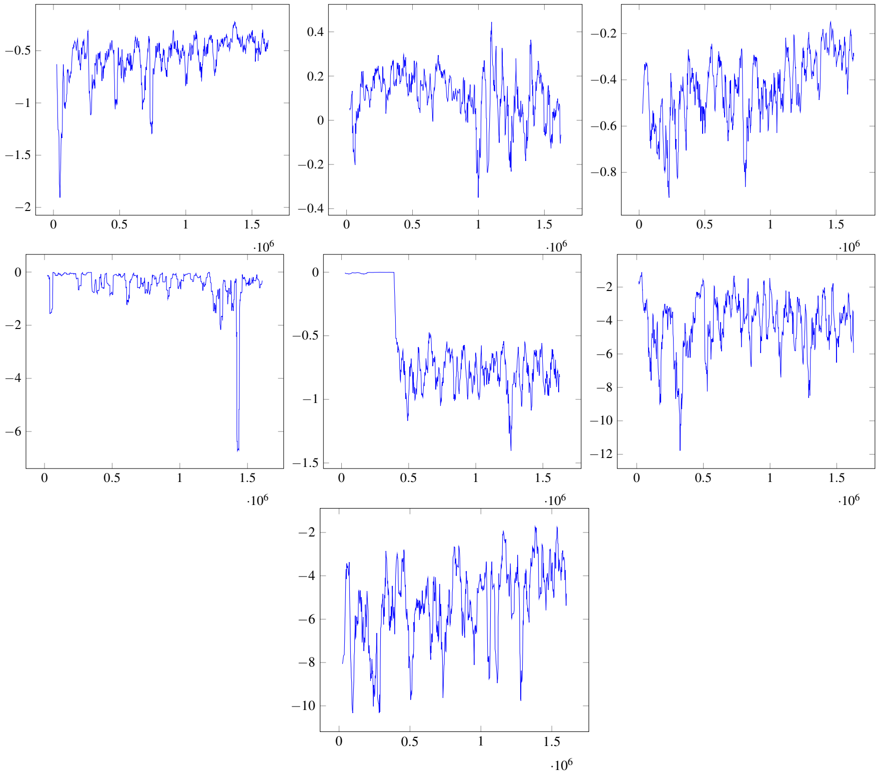

Figure 11.3 Moving average of learning curves for reward functions 1 to 7 from left to right top to bottom, for the deep model with $\rho = -1$ and $c = 0$.

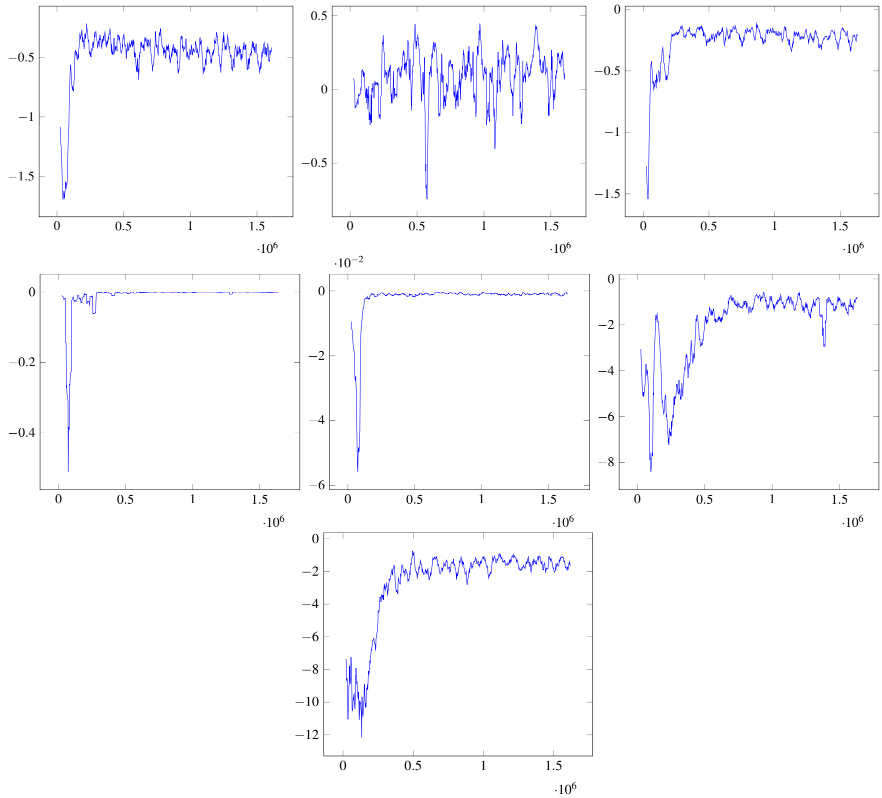

Figure 11.4 Moving average of learning curves for reward functions 1 to 7 from left to right top to bottom; for deep model with $\rho = 0$ and $c = 0$.

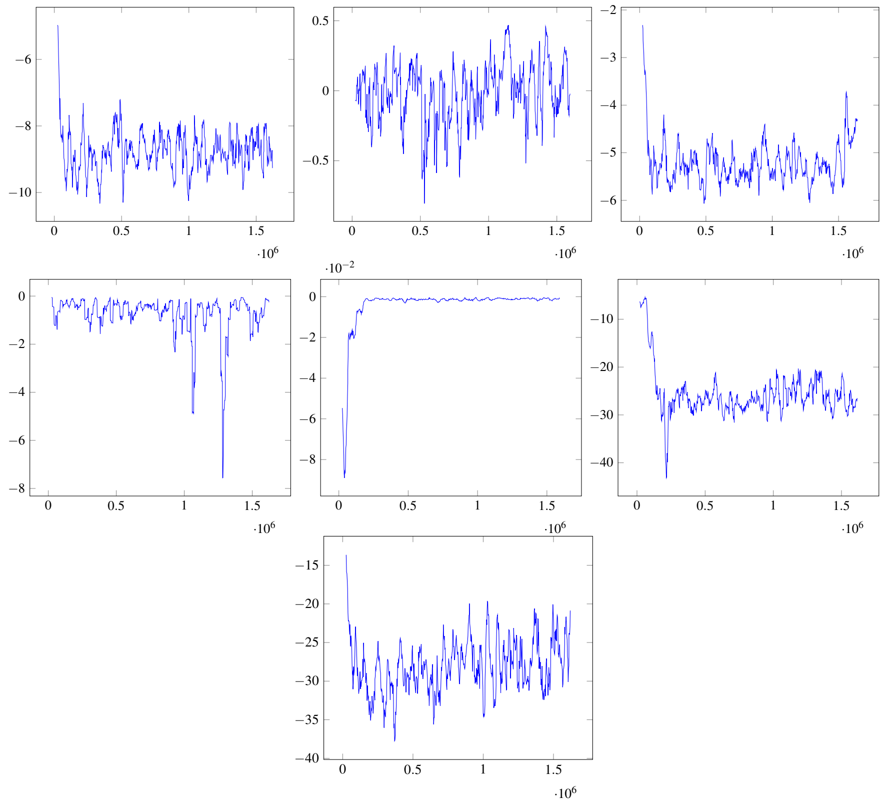

Figure 11.5 Moving average of learning curves for reward functions 1 to 7 from left to right top to bottom, for the deep model with $\rho = 0$ and $c = 0.05$.

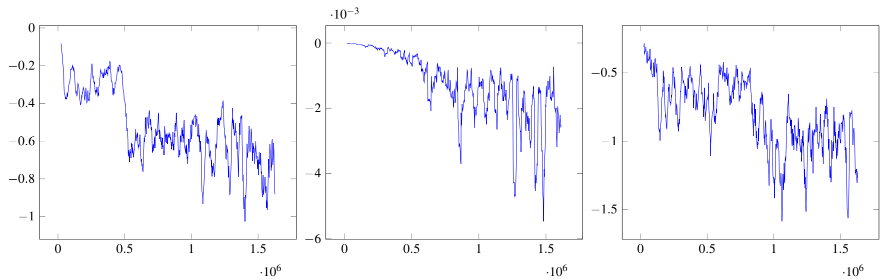

Figure 11.6 Moving average of learning curves for reward functions 1,5, and 7 respectively from left to right, for the pretrained model with $\rho = 0.5$ and $c = 0.05$.

11.4 Environment 2 figures

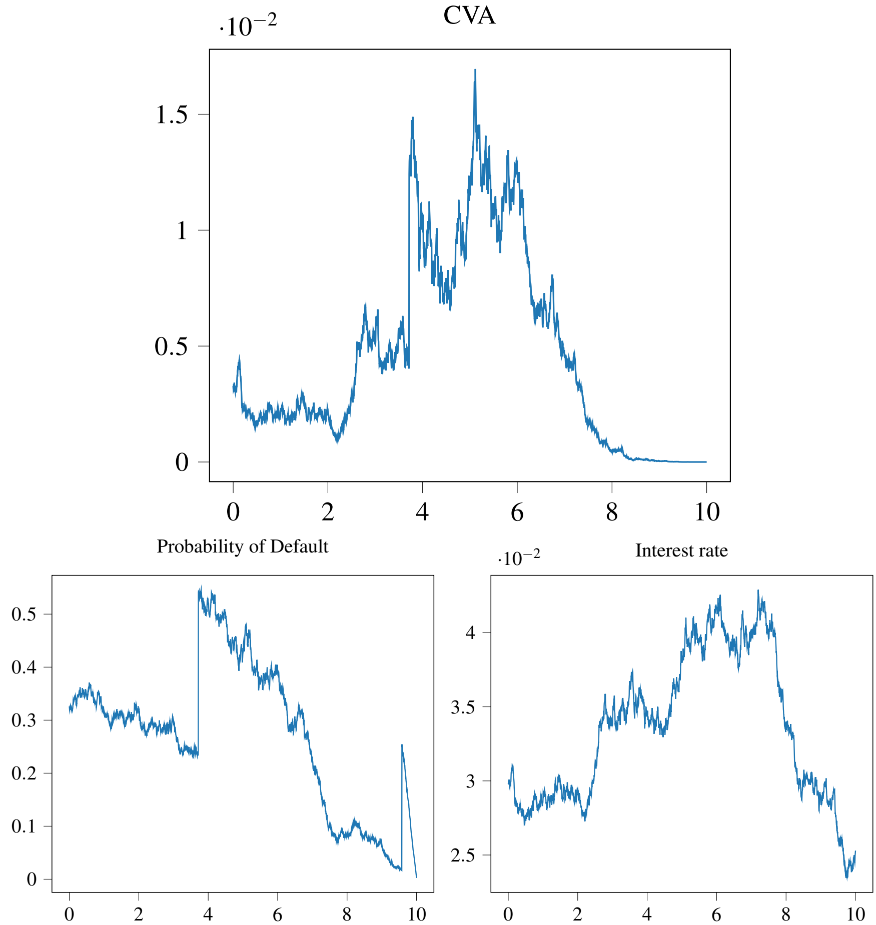

Figure 11.7 Example realization of environment 2 with no correlation.

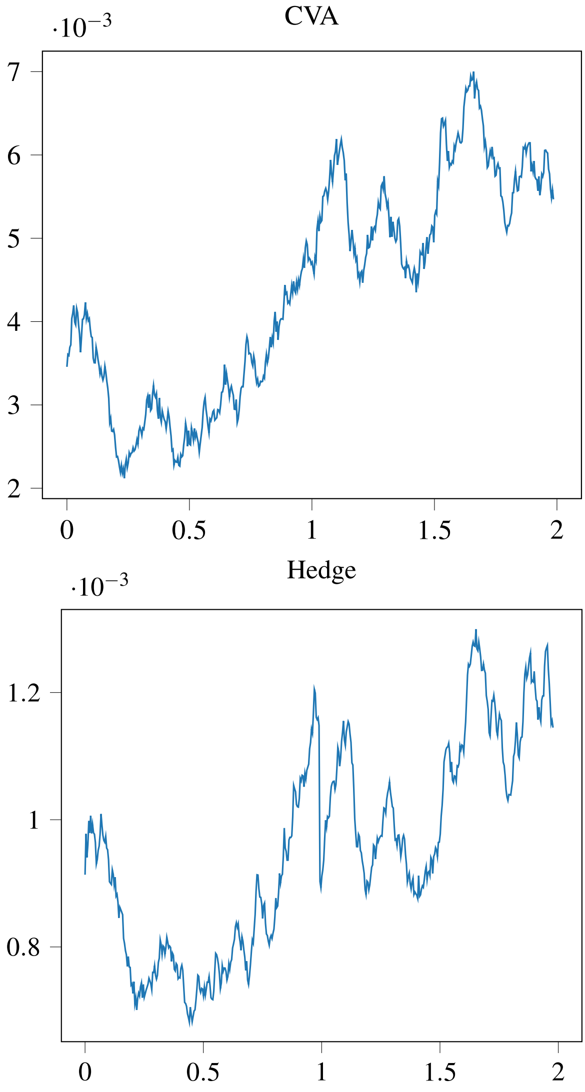

Figure 11.8 Example realization and hedging of environment 2 with -0.5 correlation and trading cost 100 percent trained with the base architecture and reward function 1.

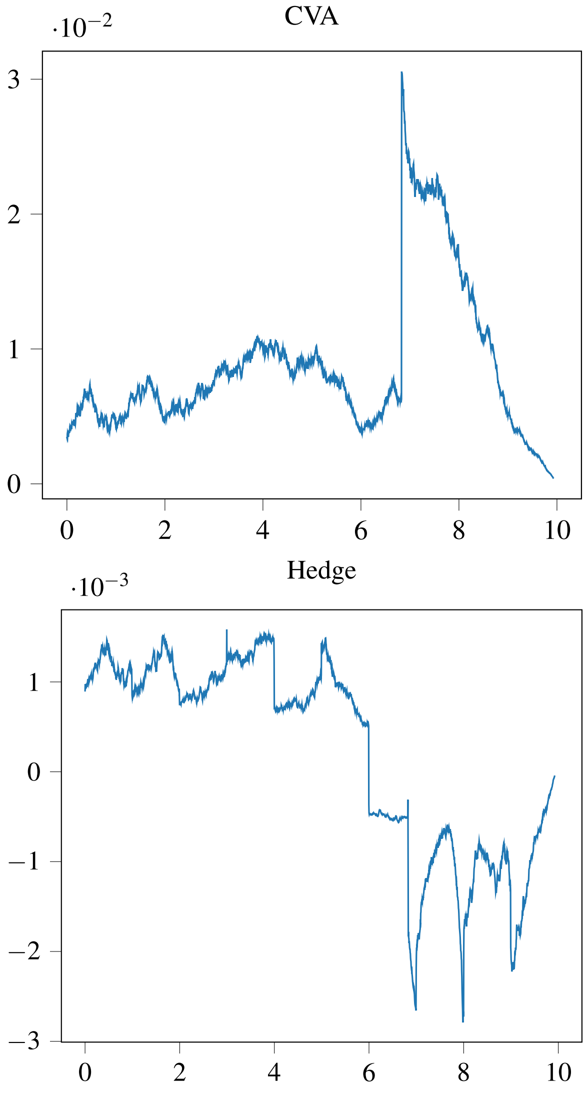

Figure 11.9 Example realization and hedging of environment 2 with -0.5 correlation and trading cost 100 percent trained with the base architecture and reward function 1.

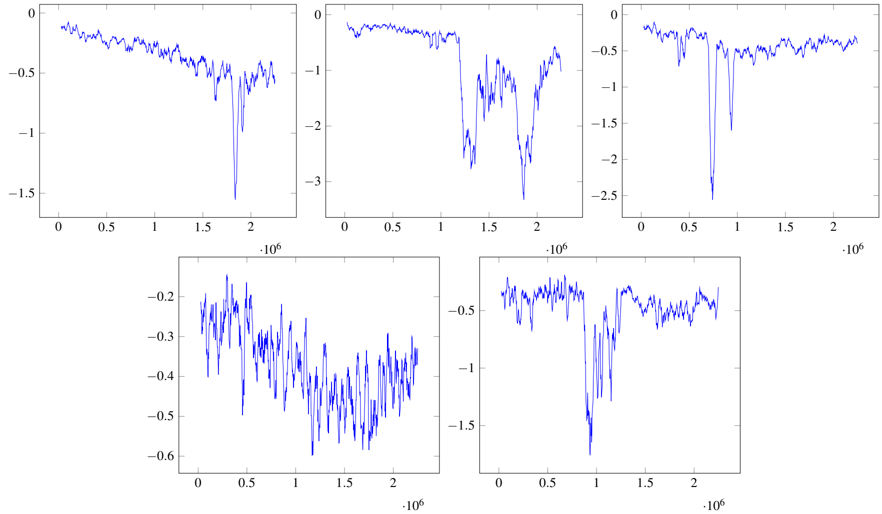

Figure 11.10 Moving average of learning curves for reward function 1 with correlations going through $\{-1, -0.5, 0, 0.5, 1\}$ from left to right top to bottom, for environment 2 with trading cost $c = 0$.

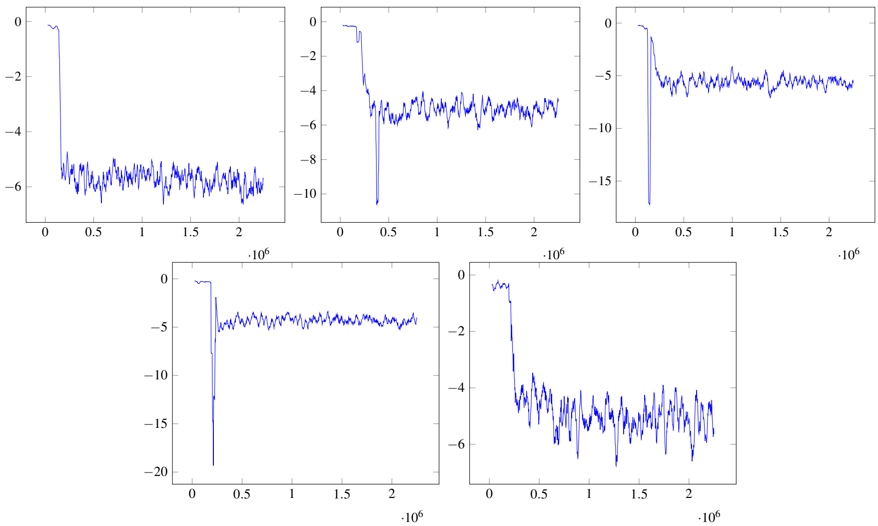

Figure 11.11 Moving average of learning curves for reward function 1 with correlations going through $\{-1, -0.5, 0, 0.5, 1\}$ from left to right top to bottom, for environment 2 with trading cost $c = 1$.

Bibliography

Ahlberg, J. (2013). *Credit value adjustment*. Master's Thesis. Lund, Sweden.

Bank for International Settlements (2020). "Targeted revisions to the credit valuation adjustment risk framework".

Bank for International Settlements (2023). *Otc derivatives statistics at end-june 2023*. Accessed: 2024-05-14. URL: https://www.bis.org/publ/otc_hy2311.htm.

Basel III: International Regulatory Framework for Banks (2021). Accessed: 2024-05-14. URL: https://www.bis.org/bcbs/basel3.htm.

Berns, C. (2014). *Cva risk charge and p&l volatility trade-off: proposal of a unified steering*. Available at SSRN: https://ssrn.com/abstract=2477164 or http://dx.doi.org/10.2139/ssrn.2477164.

Bertsekas, D. P. (2019). *Reinforcement Learning and Optimal Control*. Athena Scientific. ISBN: 9781886529397.

Bertsekas, D. P. (2023). *A Course in Reinforcement Learning*. Athena Scientific. ISBN: 9781886529496.

Björk, T. (2019). *Arbitrage Theory in Continuous Time*. Oxford University Press. ISBN: 9780198851615. DOI: 10.1093/oso/9780198851615.001.0001. URL: https://doi.org/10.1093/oso/9780198851615.001.0001.

Bomfim, A. N. (2022). "Credit default swaps". *Finance and Economics Discussion Series*. DOI: https://doi.org/10.17016/FEDS.2022.023.

Brigo, D., M. Morini, and A. Pallavicini (2013). *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases for All Asset Classes*. John Wiley & Sons Inc. ISBN: 9780470748466.

Buehler, H., L. Gonon, J. Teichmann, and B. Wood (2018). "Deep hedging". *arXiv*. DOI: 10.48550/arXiv.1802.03042. URL: https://doi.org/10.48550/arXiv.1802.03042.

Buehler, K., D. Chiarella, H. Heidegger, M. Lemerle, A. Lal, and J. Moon (2015). "The fundamental review of the trading book: implications and actions for banks". *McKinsey Working Papers on Corporate & Investment Banking | No. 11*.

Cao, J., J. Chen, J. Hull, and Z. Poulos (2021). "Deep hedging of derivatives using reinforcement learning". *The Journal of Financial Data Science* **3**. DOI: 10.3905/jfds.2020.1.052. URL: https://doi.org/10.3905/jfds.2020.1.052.

Chicot, H. (2019). "Measuring cross-gamma risk". *SSRN*. DOI: 10.2139/ssrn.3502251. URL: http://dx.doi.org/10.2139/ssrn.3502251.

Daluiso, R., M. Pinciroli, M. Trapletti, and E. Vittori (2023). "Cva hedging by risk-averse stochastic-horizon reinforcement learning". *ICAIF*.

Davis, M. H., V. G. Panas, and T. Zariphopoulou (1993). "European option pricing with transaction costs". *SIAM Journal on Control and Optimization*, pp. 470–493.

Dhariwal, P., C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, Y. Wu, and P. Zhokhov (2017). *Openai baselines*. https://github.com/openai/baselines.

Du, J., M. Jin, P. Kolm, G. Ritter, Y. Wang, and B. Zhang (2020). "Deep reinforcement learning for option replication and hedging". *The Journal of Financial Data Science* **2**, pp. 44–57. DOI: 10.3905/jfds.2020.1.045.

Esipov, S. and I. Vaysburd (1999). *On the Profit and Loss Distribution of Dynamic Hedging Strategies*. Tech. rep. 9899-03. Available at SSRN: https://ssrn.com/abstract=145172 or http://dx.doi.org/10.2139/ssrn.145172. Discussion Paper Series.

Gregory, J. (2010). *Counterparty Credit Risk - The New Challenge for Global Financial Markets*. 1st. John Wiley & Sons Ltd.

Gregory, J. (2012). *Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets*. John Wiley & Sons Inc. ISBN: 9780470748466.

Heckinger, R., I. Ruffini, and K. Wells (2014). "Understanding derivatives—markets and infrastructure: over-the-counter (otc) derivatives". *Federal Reserve Bank of Chicago*.

Hill, A., A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, and Y. Wu (2018). *Stable baselines*. https://github.com/hill-a/stable-baselines.

Hugging Face (n.d.). *Exploration vs exploitation tradeoff*. Accessed: 2024-05-23. URL: https://huggingface.co/learn/deep-rl-course/en/unit1/exp-exp-tradeoff.

Hull, J. and A. White (2012). “Cva and wrong-way risk”. *Financial Analysts Journal*.

Hull, J. C. (2018). *Options, Futures, and Other Derivatives*. 10th. Pearson.

Hull, J. C. and A. White (2017). “Optimal delta hedging for options”. *SSRN*. DOI: 10.2139/ssrn.2658343. URL: https://dx.doi.org/10.2139/ssrn.2658343.

Kolm, P. N. and G. Ritter (2019). “Dynamic replication and hedging: a reinforcement learning approach”. *The Journal of Financial Data Science*, pp. 159–171. DOI: 10.3905/jfds.2020.1.045.

Moreau, L., J. Muhle-Karbe, and H. M. Soner (2017). “Trading with small price impact”. *Mathematical Finance*, pp. 350–400.

Muhle-Karbe, J., M. Reppen, and H. M. Soner (2017). “A primer on portfolio choice with small transaction costs”. *Annual Review of Financial Economics*, pp. 301–331.

Oosterlinck, K. (2017). “History of forward contracts (historical evidence for forward contracts)”. In: *The New Palgrave Dictionary of Economics*. Palgrave Macmillan UK, London, pp. 1–3. ISBN: 978-1-349-95121-5. DOI: 10.1057/978-1-349-95121-5_2974-1. URL: https://doi.org/10.1057/978-1-349-95121-5_2974-1.

Pandit, A. (2023). “Navigating the challenges and nuances of hedging in sa-cva”. *S&P Global, Market Intelligence*. URL: https://www.spglobal.com/marketintelligence/en/mi/research-analysis/challenges-hedging-in-sacva.html.

Raffin, A., A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann (2021a). “Stable-baselines3: reliable reinforcement learning implementations”. *Journal of Machine Learning Research* **22**:268, pp. 1–8. URL: http://jmlr.org/papers/v22/20-1364.html.

Raffin, A., A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann (2021b). “Stable-baselines3: reliable reinforcement learning implementations”. *Journal of Machine Learning Research* **22**:268, pp. 1–8. URL: http://jmlr.org/papers/v22/20-1364.html.

Schulman, J., F. Wolski, P. Dhariwal, A. Radford, and O. Klimov (2017). “Proximal policy optimization algorithms”. arXiv: 1707.06347 [cs.LG].

Sutton, R. S. and A. G. Barto (2018). *Reinforcement Learning: An Introduction*. MIT Press. ISBN: 9780262039246.

Tarbert, H. P. (2020). “The enduring legacy of the dodd-frank act’s derivatives reforms”. *Journal of Financial Regulation* **6**:2, pp. 159–171. DOI: 10.1093/jfr/fjaa011. URL: https://doi.org/10.1093/jfr/fjaa011.

Zhu, S. H. and M. Pykhtin (2007). “A guide to modeling counterparty credit risk”. *SSRN*, p. 7.

Lund University
Department of Automatic Control
Box 118
SE-221 00 Lund Sweden

Document name
MASTER'S THESIS

Date of issue
June 2024

Document Number
TFRT-6253

Author(s)
Oscar Gummesson Atroshi
Osman Sibai

Supervisor
Shengyao Zhu, Nordea Markets, Lund University, Sweden
Manu Upadhyaya, Dept. of Automatic Control, Lund University
Magnus Wiktorsson, Centre for mathematical sciences, Lund University
Johan Eker, Dept. of Automatic Control, Lund University, Sweden (examiner)

Title and subtitle
Deep hedging of CVA

Abstract

Large financial institutions are vulnerable to numerous financial risks, necessitating robust regulatory frameworks to prevent crises such as those experienced in 2008. The Basel framework, devised by the Basel Committee on Banking Supervision, incorporates critical measures such as the credit valuation adjustment (CVA) to mitigate these risks. CVA fluctuates significantly based on market factors and counterparty conditions, these fluctuations need to be handled, and this is done through hedging. Hedging CVA is challenging due to its sensitivity to dynamic market conditions and the complexity of underlying assets, compounded by factors such as cross-gamma and wrong-way risk, which add significant complexity to effective risk management. This study explores the use of deep hedging, employing reinforcement learning to devise robust hedging strategies that navigate the complexities often associated with traditional analytic models. Through experimental simulations, this research compares the efficacy of traditional delta hedging with that of a reinforcement learning-based strategy, providing insights into their respective performances. The study evaluates two different market models, with the RL strategies showing promising results, particularly in the less complex model, highlighting the challenges of addressing high-dimensional problems. The findings establish a foundation for further research and demonstrate the potential of reinforcement learning in enhancing CVA hedging strategies.

Keywords

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title
0280-5316

ISBN

Language
English

Number of pages
1-106

Recipient's notes

Security classification

http://www.control.lth.se/publications/