

# Optimal Strategies for Target Volatility Contracts

Roberto Daluiso\*    Andrea Pallavicini<sup>†</sup>    Emanuele Nastasi<sup>‡</sup>  
Stefano Polo <sup>§</sup>

## Abstract

...

**JEL classification codes:** C63, G13.

**AMS classification codes:** 65C05, 91G20, 91G60.

**Keywords:** Asset Allocation, Target Volatility, Funding Costs, Hedging Cost, Reinforcement Learning, Proximal policy optimization.

---

\*Banca IMI Milan, roberto.daluiso@bancaimi.com

<sup>†</sup>Imperial College London and Banca IMI Milan, a.pallavicini@imperial.ac.uk

<sup>‡</sup>Exprivia, emanuele.nastasi@exprivia.com

<sup>§</sup>Intesa Sanpaolo Milan, stefano.polo@intesanpaolo.com

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Target Volatility Strategy</b>	<b>4</b>
2.1	The strategy Price Process . . . . .	4
2.2	The Volatility Targeting Constraint . . . . .	6
<b>3</b>	<b>Derivative Pricing</b>	<b>8</b>
3.1	European Options . . . . .	8
3.2	Stochastic Optimal Control Problem . . . . .	9
3.3	Black and Scholes Model . . . . .	10
3.3.1	Black's Closed form solution . . . . .	10
3.3.2	Active Asset (or Bang Bang) Solution . . . . .	11
<b>4</b>	<b>Reinforcement Learning</b>	<b>12</b>

# 1 Introduction

In the recent years portfolio managers are exposed to very low interest rates and quickly changing market volatilities. An effective solution to control risks under such environment is given by target volatility strategies (TVSs) (also known as constant volatility targeting) which are able to preserve the portfolio at a predetermined level of volatility. A TVS is a portfolio of risky assets (typically equities) and a risk-free asset dynamically re-balanced with the aim of maintaining the overall portfolio volatility level closed to some target value. This products were initially offered in the Asian markets, see for instance the reports of Chew, 2011 and Xue, 2012 which highlight the pros and cons for investors, to be adopted in the following years in many other markets in North America and Europe as depicted in Morrison, 2013.

In literature TVSs are tested to investigate their performances in term of realized returns, see for instance Hocquard et al., 2013 and Perchet et al., 2016, and the soundness of the volatility targeting algorithm, as described in Kim and Enke, 2018. Moreover, in pricing literature derivative contracts on TVS are studied, see in particular Di Graziano and Torricelli, 2012, Grasselli and Romo, 2016, and Albeverio et al., 2019.

In this contribution we study a problem related to TVS. We deal with the funding costs coming from hedging the risky assets underlying the TVS. We consider the point of view of a bank selling a protection to a portfolio manager on the capital invested in a TVS. the portfolio manager has the freedom of changing the relative weights of the risky assets during the life of the TVS. Since the risky assets have different hedging costs, the bank shall adjust the price of the protection to include them in the worst-case scenario. Hence, the pricing problem becomes a dynamical control problem over the risky portfolio composition. In our contribution we describe the dynamical control problem, and we derive an analytical solution in the simple case of the risky assets driven by a Black-Scholes (BS) model. Then, we tackle the problem in the general case by using reinforcement learning (RL) algorithms.

The paper is organized ad follows. In Section 2 we describe the dynamics of a TVS in presence of valuation adjustments. Then, in Section 3 we derive the analytical results, and in Section 4 we illustrate how we have applied RL to solve the dynamic control problem. We conclude the paper with ?? where we present the numerical results obtained in this work.

## 2 Target Volatility Strategy

In a TVS the fund manager selects an allocation strategy aiming at stabilizing the portfolio volatility to a target level. Clients investing in the fund pay a running fee for the service and their capital is protected. The fund manager usually buys an option on the TVS to ensure capital protection. For instance, the capital can be protected by buying a put option. In this case, we can write the net asset value (NAV)  $A_t$  of the strategy as given by<sup>1</sup>

$$A_t := \max\{V_t, K\} = V_t + (K - V_t)^+, \quad (1)$$

where  $V_t$  is the price process of the strategy, and  $K$  is the guaranteed capital. On the other hand, the fund manager can replicate the payoff by means of the put-call parity by investing the capital in a low-risk asset and buying a call on the strategy

$$A_t = K + (V_t - K)^+. \quad (2)$$

In this way the TVS is only defined in the two contracts client-fund and fund-bank. The fund manager is not implementing the strategy by trading in the market, and he is not subject to additional costs to access the market. On the other way, the bank is paying such costs since she is actively hedging the call option sold to the manager.

The bank trading activity implemented to actively hedge the option requires funding the collateral procedures of the hedging instruments along with any lending/borrowing fee. The price of a financial product sold by the bank is modified to include any valuation adjustment due to the trading activity. We proceed by defining the price process for the TVS so that we can highlight the impact of valuation adjustments.

### 2.1 The strategy Price Process

We work on a filtered probability space  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$  satisfying the usual assumptions for a market model, where  $\mathbb{P}$  is the physical probability measure representing the actual distribution of supply and demand shocks on equities prices.

We consider a fund trading a basket of  $n$  risky securities with price process  $S_t^i$  with  $i = 1, \dots, n$  funded with a cash account  $B_t$  accruing at  $r_t$ . Any dividend paid by the securities is re-invested in the fund, so that we limit our analysis to total return securities, namely we assume that holding the security is self-financing. Here, we assume that the TVS is implemented

---

<sup>1</sup>Here we neglect discounting factors.

in continuous time, even if in the practice we can implement the strategy only on a discrete set of dates. We introduce the deflated gain process  $\bar{G}_t^i$  associated to the risky securities as given by

$$\bar{G}_t^i := \bar{S}_t^i + \bar{D}_t^i, \quad (3)$$

where we define the deflated price<sup>2</sup> and cumulative dividend processes as

$$\bar{S}_t^i := \frac{S_t^i}{B_t}, \quad \bar{D}_t^i := \int_0^t \frac{d\pi_u^i}{B_u} + \int_0^t \frac{d\psi_u^i}{B_u}, \quad (4)$$

where  $\pi_t^i$  represents the cumulative contractual-coupon process paid by the security, and  $\psi_t^i$  represents the cumulative valuation adjustments.

Valuation adjustments (XVAs) is a topic widely discussed in the literature. We refer to Brigo et al., 2013 for a discussion. Since fund managers allocating TVS usually rely on Equity assets, here we use the results of Gabrielli et al., 2020 which analyze the valuation adjustments for equity products. We can write

$$\psi_t^i := \int_0^t S_u^i \mu_u^i du, \quad (5)$$

where we call  $\mu_t^i$  cost of carry, which basically represents the hedging costs for the  $i$ -th security.

Then, we introduce the strategy price process  $V_t$ , and we define the deflated gain process  $\bar{G}_t^V$  as given by

$$\bar{G}_t^V := \frac{V_t}{B_t} + \int_0^t \frac{V_u \phi_u}{B_u} du, \quad (6)$$

where  $\phi_t$  are the running fees earned by the fund manager for his activity. We assume that the strategy is self-financing, so that we can write

$$d\bar{G}_t^V = q_t \cdot d\bar{G}_t, \quad (7)$$

where  $q_t^i$  is the quantity invested in the  $i$ -th security<sup>3</sup>.

Now, in order to prevent arbitrages, we assume the existence of a risk-neutral measure  $\mathbb{Q}$  under which the deflated gain processes of all traded security are martingales. Under this assumption we are able to derive the

<sup>2</sup>We use bar notation for deflated quantities: processes expressed in terms of  $B_t$ .

<sup>3</sup>In all formulae we use dot notation for scalar product between vectors, i.e.  $a \cdot b = \sum_i a_i b_i$ , or between matrix and vector, i.e.  $A \cdot b = \sum_j a_{ij} b_j$  or  $b \cdot A = \sum_i b_i a_{ij}$ .

drift conditions on the security price processes, and in turn on the strategy price process.

$$\forall T > t \quad \bar{G}_t^i = \mathbb{E}_t [\bar{G}_T^i] \implies dS_t^i = r_t S_t^i dt - d\pi_t^i - d\psi_t^i + dM_t^i, \quad (8)$$

where  $M_t^i$  are martingale under  $\mathbb{Q}$ . If we substitute this expression for the security dynamics into the definition of the strategy we can check that the price process of the strategy is accruing at a cash account rate rate  $r_t$  compensated for the fund manager fees

$$dV_t = V_t(r_t - \phi_t)dt + dM_t^V, \quad (9)$$

with  $M_t^V$  martingale under  $\mathbb{Q}$ . Notice that, as expected from non-arbitrage considerations, the coupons paid by each security appear only in the drift of the security price process, but they do not impact the drift of the strategy.

Yet, the strategy priced by  $V_t$  cannot be described in the contract between the parties, since Equation (7) depends via the security gain processes on the valuation adjustment  $\psi_t^i$ , which is specific of the investor pricing the strategy. Thus, the TVS defined in the contract will be

$$d\bar{I}_t := q_t \cdot \left( d\bar{S}_t + \frac{d\pi_t}{B_t} \right) - \bar{I}_t \phi_t dt \quad \text{with } I_0 = V_0, \quad (10)$$

leading to the following price process dynamics

$$dI_t = I_t(r_t - \phi_t)dt - q_t \cdot \psi_t + dM_t^I, \quad (11)$$

with  $M_t^I$  martingale under  $\mathbb{Q}$ . In this case we observe that  $I_t$  depends explicitly both on the valuation adjustments and on the allocation strategy. Indeed, if we substitute the valuation adjustments with their explicit expression (Equation (5)), we get

$$dI_t = I_t(r_t - \phi_t)dt - q_t \cdot S_t \mu_t dt + dM_t^I, \quad (12)$$

where we can see the dependency on cost of carry  $\mu_t^i$ .

## 2.2 The Volatility Targeting Constraint

In a typical TVS the fund manager selects a risky-asset portfolio with a specific time-dependent allocation strategy expressed by means of the vector of relative weights  $\alpha_t$ , along with a risk-free asset, which we can identify with

the bank account  $B_t$ . Usually TVSs are total-return products; thus we are justified in assuming  $\pi_t = 0$ . Thus we can write Equation (10) as given by

$$\frac{dI_t}{I_t} = \omega_t \alpha_t \cdot \frac{dS_t}{S_t} + (1 - \omega_t \alpha_t \cdot \mathbf{1}) \frac{dB_t}{B_t} - \phi_t dt, \quad (13)$$

where  $\mathbf{1}$  is a  $n$ -dimensional vector of ones and  $\omega_t \in [0, 1]$  is determined so that the strategy log-normal volatility is kept constant, namely

$$\omega_t : \quad \text{Var}_t[dI_t] = \bar{\sigma}^2 I_t^2 dt, \quad (14)$$

where  $\bar{\sigma}$  is the target volatility value.

We assume a generic continuous semi-martingale dynamics under the risk-neutral measure for the underlying securities, so that we can write In practice, this means that the fund manager will select a risky-portfolio choosing  $\alpha_t$  equities from the universe where he can trade and after that his choices will be scaled by the automatic target volatility algorithm<sup>4</sup>  $\omega_t$ .

To derive the expression for  $\omega_t$  we need to assume a generic continuous semi-martingales dynamics under the risk-neutral measure for the underlying securities, so that we can write Equation (8) as

$$\frac{dS_t^i}{S_t^i} = (r_t - \mu_t^i) dt + \nu_t^i \cdot dW_t, \quad (15)$$

where  $\nu_t$  is an adapted matrix process ensuring the existence of a solution for the SDE and  $W_t$  is a  $n$ -dimensional vector of Brownian motions under  $\mathbb{Q}$ . Under these assumptions we can derive an expression for  $\omega_t$ , and we get<sup>5</sup>

$$\omega_t = \frac{\bar{\sigma}}{\|\alpha_t \cdot \nu_t\|}. \quad (16)$$

Hence, putting this last result in the dynamics of  $I_t$  we obtain

$$\frac{dI_t}{I_t} = \left( r_t - \phi_t - \frac{\bar{\sigma} \alpha_t}{\|\alpha_t \cdot \nu_t\|} \cdot \mu_t \right) dt + \frac{\bar{\sigma} \alpha_t}{\|\alpha_t \cdot \nu_t\|} \cdot \nu_t \cdot dW_t, \quad (17)$$

where we can see, as expected, that the strategy grows at the risk-free rate but for adjustments due to valuation adjustments and fees.

<sup>4</sup>We recall that the universe of assets where the manager can trade and the value of  $\bar{\sigma}$  are written in the contract.

<sup>5</sup>In all formulae the norm for a vector  $a$  is defined as  $\|a\| := \sqrt{a \cdot a}$ .

### 3 Derivative Pricing

A derivative contract on the TVS with maturity  $T$  can be defined as

$$V_0 := \sup_{\alpha} \mathbb{E}_0 \left[ \int_0^T D(0, u; \zeta) d\pi_u(\alpha) \right], \quad (18)$$

where  $D(0, T; \zeta)$  is the discount factor with rate  $\zeta_t$ , inclusive of the derivative valuation adjustments, and  $\pi_t$  is the cumulative coupon process paid by the derivative, and it depends on the allocation strategy since in turn the TVS depends on it via the valuation adjustments. We take the supremum over the strategies since we do not have any information on the future activity of the fund manager.

#### 3.1 European Options

If the derivative contract depends only on the marginal distribution of  $I_t$  at maturity (a European payoff), we are able to prove that exists an optimal strategy, and we are able to calculate it. We consider the following pricing problem

$$V_0 := \sup_{\alpha} \mathbb{E}_0 [D(0, T; \zeta) \Phi(I_t(\alpha))], \quad (19)$$

where  $\Phi$  is the payoff function of the derivative. We start by introducing the Markovian projection of the dynamics followed by  $I_t$ . We name it  $I_t^{\text{MP}}$ , and we get by applying the Gyöngy Lemma

$$\frac{dI_t^{\text{MP}}}{I_t^{\text{MP}}} := \left( r_t - \ell(t, I_t^{\text{MP}}) \right) dt + \bar{\sigma} dW_t^{\text{MP}}, \quad (20)$$

where the local drift is defined as

$$\ell(t, K) := \bar{\sigma} \mathbb{E}_0 \left[ \frac{\mu_t \cdot \alpha_t}{\|\alpha_t \cdot \nu_t\|} \middle| I_t = K \right], \quad (21)$$

and  $W_t^{\text{MP}}$  is a Brownian motion under the risk-neutral measure. Notice that the diffusion coefficient collapses to the target volatility value  $\bar{\sigma}$ . Since European payoffs depends only on the marginal distribution at maturity, they can be calculated by means of the Markovian projection  $I_t^{\text{MP}}$ , namely

$$V_0 := \sup_{\alpha} \mathbb{E}_0 \left[ D_0(T) \Phi(I_T^{\text{MP}}(\alpha)) \right]. \quad (22)$$

Hence, we have our first result valid only if valuation adjustments can be neglected:



**Proposition 3.1.** *A European payoff on the TVS can be calculated by assuming any allocation in the underlying risky basket if all the underlying securities grow under the risk-neutral measure at the risk-free rate without any valuation adjustment, namely if we can write  $\mu_t = 0$ .*

**Remark 3.1** (Existence of the Solution in the General Case). *In a more general settings we are not able to find an explicit solution. A proof of the existence of the solution in a general setting is missing. This is a stochastic optimal control problem where by homogeneity we can suppose that  $\alpha_t$  lives in a compact domain, namely (a subset of) the unit simplex. At least if  $r_t$ ,  $\mu_t^i$  and  $\nu_t^i$  are (uniformly) bounded, and if the eigenvalues of  $\nu_t$  are (uniformly) bounded away from zero, then drift and diffusion should be uniformly Lipschitz in  $\alpha_t$ , and classical theorems should exist.*

### 3.2 Stochastic Optimal Control Problem

In presence of valuation adjustments we need to solve the full optimization problem. We discretize the optimal strategy  $\alpha_t$  as

$$\alpha_t := \sum_k \mathbf{1}_{\{t \in [T_{k-1}, T_k)\}} \alpha_{T_{k-1}}, \quad (23)$$

according to a time grid  $\mathcal{T} := \{T_0, \dots, T_k, \dots, T_m\}$  with  $T_0 := t$  the pricing date and  $T_m := T$  the maturity of the option. Therefore we can apply the dynamic programming principle to express the optimal  $\alpha_t$  at time  $T_{k-1}$  as

$$\alpha_{T_{k-1}} := \arg \max_{\alpha} \left\{ \mathbb{E}_{T_{k-1}} \left[ D_{T_{k-1}}(T_k) V_{T_k}(X_{T_k}, I_{T_k}(\alpha)) \mid X_{T_{k-1}}, I_{T_{k-1}} \right] \right\}, \quad (24)$$

where  $V_{T_k}$  is the option value at time  $T_k$  and  $X$  is any Markovian state such that the drift and the diffusion coefficient of  $I_t$  are a function of  $(X_t, I_t, \alpha_t)$ . We calculate  $I_{T_k}(\alpha_{T_{k-1}})$  for any given strategy  $\alpha_{T_{k-1}}$  by a suitable discretization of (17) starting from  $X_{T_{k-1}}$  and  $I_{T_{k-1}}$ .

Thus the derivative price is given by:

$$V_{T_{k-1}}(X_{T_{k-1}}, I_{T_{k-1}}) = \mathbb{E}_{T_{k-1}} \left[ D_{T_{k-1}}(T_k) V_{T_k}(X_{T_k}, I_{T_k}(\alpha_{T_{k-1}})) \mid X_{T_{k-1}}, I_{T_{k-1}} \right], \quad (25)$$

while the iteration starts from maturity date where the boundary condition is set equal to the payoff function:

$$V_{T_m} = \Phi(I_{T_m}). \quad (26)$$

### 3.3 Black and Scholes Model

In the Black and Scholes model with deterministic rates, we can work with empty  $X_t$ , since in this case the portfolio dynamics (17) is Markovian, leading to an optimal strategy  $\alpha_t^*$  which depends in principle only on  $I_t$ . As a consequence, the local drift can be written as

$$\ell(t, K) = \bar{\sigma} \frac{\mu(t) \cdot \alpha(t, K)}{\|\alpha(t, K) \cdot \nu(t)\|}, \quad (27)$$

so that the optimization problem can be solved looking only at the Markovian projection without simulating all the Brownian motions  $W_t$ . Notice that we are indicating the dependency on time in parenthesis to highlight that in this formula all the quantities are deterministic function of time.

A direct consequence is the following proposition, which is relevant for plain vanilla options on TVS.

**Proposition 3.2.** *When the underlying securities follow a Black and Scholes model with deterministic rates, the optimal strategy for a monotone payoff consists in minimizing the local drift function, independently of the current state  $I_t$*

$$\alpha^*(t) := \arg \min_{\alpha} \frac{\alpha \cdot \mu(t)}{\|\alpha \cdot \nu(t)\|}. \quad (28)$$

The absence of stochastic elements in Equation (28) makes the optimal strategy known *a priori*; in fact one can solve the optimization problem once for all for each  $t \in \mathcal{T}$  just looking at the market data  $\mu(t)$  and  $\nu(t)$  for the securities. Once  $\alpha^*$  is known, then one can price the payoff by Monte Carlo simulation.

#### 3.3.1 Black's Closed form solution

In absence of constraints on the allocation strategy, we are able to derive a closed form solution to the BS problem (28)

**Lemma 3.3.** *Let be  $\mu, \alpha \in \mathbb{R}^n$ ,  $\nu \in \mathbb{R}^{n \times n}$  be a full rank matrix and  $\Sigma := \nu \nu^T$ . Then the closed solution of the optimization problem (28) is*

$$\alpha^* = - \frac{\Sigma^{-1} \cdot \mu}{\|(\Sigma^{-1} \cdot \mu) \cdot \nu\|} \quad (29)$$

*Proof.* Since the argument of the minimum (28) is zero-homogeneous, then we can rewrite the problem as

$$\begin{aligned} & \text{minimize } \alpha \cdot \mu \\ & \text{subject to } \|\alpha \cdot \nu\|^2 = 1 \end{aligned} \quad (30)$$

By setting the Lagrangian function associated with the problem

$$\mathcal{L}(\alpha, \lambda) = \alpha \cdot \mu - \lambda (\|\alpha \cdot \nu\|^2 - 1), \quad (31)$$

we obtain the first order conditions

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \alpha} = \mu - 2\lambda \Sigma \cdot \alpha = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = \|\alpha \cdot \nu\|^2 - 1 = 0, \end{cases} \quad (32)$$

Then, by applying simple algebra, we obtain the analytical form of the free optimal strategy

$$\alpha^* = \pm \frac{\Sigma^{-1} \cdot \mu}{\|(\Sigma^{-1} \cdot \mu) \cdot \nu\|} \quad (33)$$

We take the minus sign to get the minimum value of the TVS local drift.  $\square$

### 3.3.2 Active Asset (or Bang Bang) Solution

A closed form solution to the minimization of the local drift correction (28) can also be derived in the common case that all costs of carry are nonnegative and the only constraint on portfolio weights is nonnegativity, which would mean a long only strategy by the fund manager.

**Lemma 3.4.** *Let  $\mu \in \mathbb{R}^n$  be a vector with nonnegative components,  $\nu \in \mathbb{R}^{n \times n}$  be a full rank matrix, and  $\Sigma = \nu \nu^T$ . Then*

$$\inf_{\alpha \in \mathbb{R}_+^n \setminus \{0\}} \frac{\alpha \cdot \mu}{\|\alpha \cdot \nu\|} = \min_{i \leq n} \frac{\mu_i}{\sqrt{\Sigma_{ii}}}; \quad (34)$$

*if  $\bar{i}$  is the index which realizes the min, then the infimum is realized by a vector concentrated on the  $\bar{i}$  component:  $\alpha_i = \delta_{i\bar{i}}$ .*

*Proof.* Let us first consider the case in which  $\mu = \mathbb{1}$ . Since the argument of the infimum is zero-homogeneous, normalizing by  $\alpha \cdot \mathbb{1} > 0$  we can restrict to the affine hyperspace  $\{\alpha \cdot \mathbb{1} = 1\}$ , where the minimization (34) reduces to the maximization of its denominator: the required infimum will be square root of the reciprocal of

$$\sup \{ \|\alpha \cdot \nu\|^2 \mid \alpha \in \mathbb{R}_+^n, \alpha \cdot \mathbb{1} = 1 \}.$$

Now we can note that  $\Sigma$  is positive definite, hence  $\Sigma_{ij} < \sqrt{\Sigma_{ii}\Sigma_{jj}} \leq \Sigma_{\bar{i}\bar{i}}$ , which implies

$$\|\alpha \cdot \nu\|^2 = \sum_{i,j=1}^n \alpha_i \alpha_j \Sigma_{ij} \leq \sum_{i,j=1}^n \alpha_i \alpha_j \Sigma_{\bar{i}\bar{i}} = \Sigma_{\bar{i}\bar{i}}$$

because  $\sum_i \alpha_i = 1$ . Since we trivially have equality for  $\alpha_i = \delta_{ii}$ , this concludes the proof of the case  $\mu = \mathbb{1}$ .

Next, let us consider the case in which all components of  $\mu$  are strictly positive, and define  $M$  as the diagonal matrix with diagonal  $\mu$ . Then we can rewrite the infimum as a function of  $\beta = M\alpha$ :

$$\inf_{\beta \in \mathbb{R}_+^n \setminus \{0\}} \frac{\beta \cdot \mathbb{1}}{\|\beta \cdot M^{-1}\nu\|},$$

which by the first part of the proof equals

$$\min_{i \leq n} \frac{1}{\sqrt{\Sigma_{ii}}} = \min_{i \leq n} \frac{\mu_i}{\sqrt{\Sigma_{ii}}}, \quad \tilde{\Sigma} := M^{-1}\nu\nu^T M^{-1} = M^{-1}\Sigma M^{-1}.$$

Finally, let us consider the general case in which  $\mu$  may have some components equal to zero. For an arbitrary  $\epsilon \geq 0$  let us define

$$f_\epsilon(\alpha) = \frac{\alpha \cdot (\mu + \epsilon)}{\|\alpha \cdot \nu\|}.$$

One can easily note that as  $\epsilon \rightarrow 0$ ,  $f_\epsilon$  tends to  $f_0$  uniformly on the compact set  $\{\alpha \in \mathbb{R}_+^n \mid \alpha \cdot \mathbb{1} = 1\}$ , so that the minimum converges to the minimum on that set. Since we know by homogeneity that the minimum on  $\{\alpha \in \mathbb{R}_+^n \mid \alpha \cdot \mathbb{1} = 1\}$  equals the minimum on  $\mathbb{R}_+^n \setminus \{0\}$ , we conclude

$$\inf_{\alpha \in \mathbb{R}_+^n \setminus \{0\}} f_0(\alpha) = \lim_{\epsilon \rightarrow 0+} \inf_{\alpha \in \mathbb{R}_+^n \setminus \{0\}} f_\epsilon(\alpha) = \lim_{\epsilon \rightarrow 0+} \min_{i \leq n} \frac{\mu_i + \epsilon}{\sqrt{\Sigma_{ii}}} = \min_{i \leq n} \frac{\mu_i}{\sqrt{\Sigma_{ii}}}.$$

□

## 4 Reinforcement Learning

Reinforcement Learning (RL) describes how an agent behaves in an environment so to maximize some notion of cumulative reward. The actions of the agent as a function of his observations of the environment are termed the agent policy. In our case the policy is the asset allocation weights  $\alpha$ , while the rewards are the costs generated by hedging a derivative contract on the TVS. Once the agent is trained, and the optimal policy is selected, we can run a Monte Carlo simulation to calculate the option price on the TVS.

We adopt as learning strategy the PPO algorithm developed in Schulman, Wolski, et al., 2017 and Schulman, Moritz, et al., 2015. We have

## References

- Albeverio, S., S. Victoria, and K. Wallbaum (July 2019). “The volatility target effect in investment-linked products with embedded American-type derivatives”. In: *Investment Management and Financial Innovations* 16, pp. 18–28.
- Brigo, Damiano, Massimo Morini, and Andrea Pallavicini (Apr. 2013). *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases For All Asset Classes*. Wiley, Chichester.
- Chew, Yuhong (Apr. 2011). *Target Volatility Asset Allocation Strategy*. Tech. rep. Society of Actuaries.
- Di Graziano, Giuseppe and Lorenzo Torricelli (2012). “Target Volatility Option Pricing”. In: *International Journal of Theoretical and Applied Finance* 15.01.
- Gabrielli, Stefania, Andrea Pallavicini, and Stefano Scoleri (June 2020). “Funding Adjustments in Equity Linear Products”. In: *Risk*.
- Grasselli, Martino and Jacinto Marabel Romo (2016). “Stochastic Skew and Target Volatility Options”. In: *Journal of Futures Markets* 36.2, pp. 174–193.
- Hocquard, Alexandre, Sunny Ng, and Nicolas Papageorgiou (2013). “A Constant-Volatility Framework for Managing Tail Risk”. In: *The Journal of Portfolio Management* 39.2, pp. 28–40.
- Kim, Youngmin and David Enke (2018). “A dynamic target volatility strategy for asset allocation using artificial neural networks”. In: *The Engineering Economist* 63.4, pp. 273–290.
- Morrison Steven; Tadrowski, Laura (Sept. 2013). *Guarantees and Target Volatility Funds*. Tech. rep. Moody’s Analytics.
- Perchet, Romain et al. (Jan. 2016). “Predicting the Success of Volatility Targeting Strategies: Application to Equities and Other Asset Classes”. In: *The Journal of Alternative Investments* 18, pp. 21–38.
- Schulman, John, Philipp Moritz, et al. (June 2015). “High-Dimensional Continuous Control Using Generalized Advantage Estimation”. In: *arXiv preprint arXiv:1506.02438*.
- Schulman, John, Filip Wolski, et al. (2017). “Proximal Policy Optimization Algorithms”. In: *CoRR* abs/1707.06347. arXiv: 1707.06347. URL: <http://arxiv.org/abs/1707.06347>.
- Xue, Yuhong (Oct. 2012). *Target Volatility Fund: An Effective Risk Management Tool for VA?* Tech. rep. Society of Actuaries.