Reinforcement Learning for DVA Hedging

Advisor:
Marcello RESTELLI
Co-advisor:
Matteo PIROTTA

Thesis presentation:
Giorgio VIT

Motivations & Objectives

"...the own credit risk must be contemplated into the fair value measurement of a derivative."

- IFRS 13 Fair Value Measurement.

- *Can we use Neural Networks to extract patterns from financial series?*
- *Can we construct an algorithm to hedge the DVA which beats experienced human traders?*

Index

1. **DVA Hedging**: DVA formal definition, problem formalization
2. **Reinforcement Learning**: agent-environment interaction, Natural Policy Gradient
3. **Results**: dataset, P&L optimization, efficient frontier

DVA

Debt Value Adjustment (DVA)

The DVA is the risk that the bank defaults and does not pay the derivative to the client. At time t, the DVA is given by:

$$\mathrm{DVA}(t) = \mathbb{E}_t^Q[\mathrm{LGD}_B \, 1_{\{\tau_B \leq T\}} \, 1_{\{\tau_B < \tau_C\}} \, D(t, \tau_B) \, (V_0(\tau_B))^-]$$

where:

- LGD_B is the Loss Given Default of the bank.
- $\tau_{B/C}$ is the time to default (B bank, C counterparty).
- $D(t, \tau_B)$ is the risk-free stochastic discount factor evaluated in τ_B.
- $V_0(\tau_B))^-$ is the negative part of the derivative's value at the investor time to default.

Simplified Version For A First Implementation

DVA generated by the liability represented by a single cash flow N that the bank must pay at time T (5Y rolling) $\Rightarrow V_0(\tau_B) = N\, D(\tau_B, T)$

$$DVA(t) = N\, LGD_B\, \mathbb{E}_t[1_{\{\tau_B \leq T\}}\, D(t, \tau_B)\, D(\tau_B, T)]$$

Hyp:

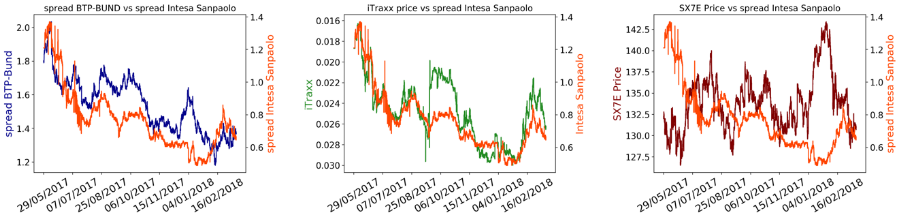
1. Independence between $1_{\{\tau_B \leq T\}}$ and $D(t, T)$
2. $r_{RiskFree} = 0$
3. Jarrow and Turnbull model for survival probability

$$\boxed{DVA(t) = N \cdot LGD_B \cdot \left(1 - e^{-\frac{\pi_t^{5y}}{LGD_B}(T-t)}\right)}$$

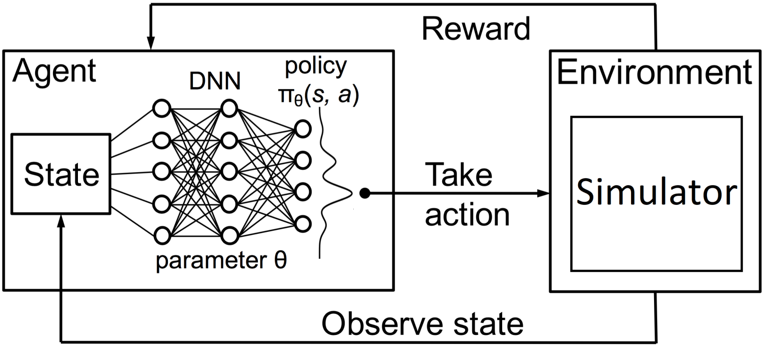
DVA Hedging

Possible trades:

1. *BTP Spread Trade:* purchase/sale of 10y BTP futures and simultaneous sale/purchase of 10y Bund Futures.
2. 5y iTraxx Financial Senior (FinSen) CDS index.
3. Futures on the Eurostoxx Banks SX7E.



Agent-Environment Interaction



Policy - Objective Function

Policy

A policy is a function $\pi : \mathbb{S} \times \mathbb{A} \to \mathbb{R}$ *such that for every* $s \in \mathbb{S}$, $A \in \mathcal{A} \to \pi(s, A)$ *is a probability distribution over* $(\mathbb{A}, \mathcal{A})$.

Objective Function

The objective function $J(\theta)$ is the expected reward that can be achieved starting from a random initial state and following the policy π_θ. In an episodic environment:

$$J(\theta) = \mathbb{E}_{\pi_\theta}[\sum_{k=0}^{H} \gamma^k r_k]$$

Natural Policy Optimization (NPO)

input: initial policy parameterization θ_0.
return: optimal policy parameters $\theta^* = \theta_{m+1}$.

while *policy parameterization* $\theta_m \approx \theta_{m+1}$ *converges* **do**
  obtain policy gradient $\nabla J(\theta_m)$ from estimator
  update policy $\theta_{m+1} = \theta_m + \alpha_m \mathbf{F}_\theta^{-1} \nabla J(\theta_m)$
end

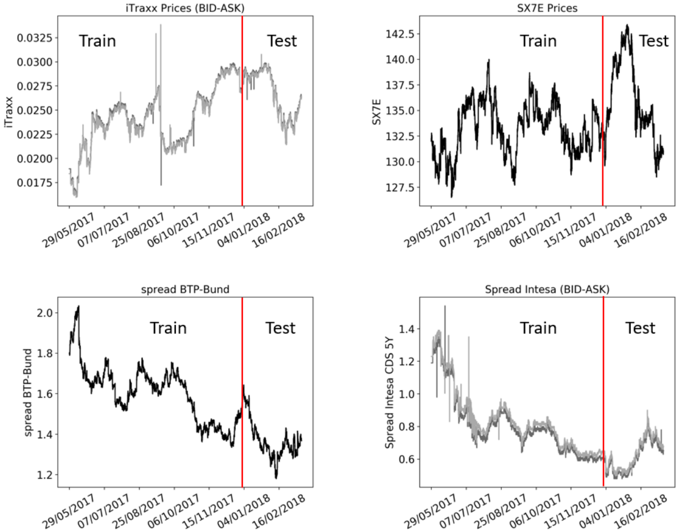
$$\nabla J(\theta) = \int_{\mathcal{T}} \nabla_\theta \mathcal{P}_\theta(\tau)\, r(\tau)\, d\tau = \mathbb{E}[\nabla_\theta log \mathcal{P}_\theta(\tau)\, r(\tau)] \approx \left\langle \sum_{k=0}^{H} \left(\sum_{l=k}^{H} \nabla_\theta log \pi_\theta(A_k|s_k) \right) (r_l - b) \right\rangle$$

$$\mathbf{F}_\theta = \mathbb{E}_{d^\pi(s)} \left[\mathbb{E}_{\pi(a;s,\theta)} \left[\frac{\partial log \pi(a;s,\theta)}{\partial \theta_i} \frac{\partial log \pi(a;s,\theta)}{\partial \theta_j} \right] \right] \approx \left\langle \sum_{k=0}^{H} \left(\sum_{l=0}^{k} \nabla_\theta log \pi_\theta(A_l|s_l) \right) \nabla_\theta log \pi_\theta(A_k|s_k)^{\mathbf{T}} \right\rangle$$

Algorithm: General setup for NPO.

Dataset

n records per day = 96
n days train = 130
n days test = 43
n records train = 12480
n records test = 4128



State

1. Intesa Sanpaolo spread CDS 5Y: π_i^{5y}
2. DVA: DVA_i; total sensitivity ($\psi_i^{\text{DVA}} \times D_i^{\text{DVA}}/10^4$) to the Intesa CDS
3. iTraxx (price: X_i^{iTraxx}, sensitivity: d_i^{iTraxx}, allocation: L_i^{iTraxx}, delta allocation: ΔL_i^{iTraxx}, spread)
4. BTP (price: X_i^{BTP}, sensitivity: d_i^{BTP}, allocation: L_i^{BTP}, delta allocation: ΔL_i^{BTP})
5. Bund (price: X_i^{Bund}, sensitivity: d_i^{Bund}, allocation: L_i^{Bund}, delta allocation: ΔL_i^{Bund})
6. BTP-Bund yield spread: $s_i^{\text{BTP}-Bund}$
7. SX7E (price:X_i^{SX7E}, sensitivity:d_i^{SX7E}, allocation:L_i^{SX7E}, delta allocation:ΔL_i^{SX7E})
8. Time to roll for differentes instruments
9. Regulatory Capital (RC).
10. VIX Index and V2X Index

Baseline

Baseline

A baseline is a simple strategy that is used to measure the performance of our RL agent's policy.

BTP Baseline

$$\begin{cases} A_{k+1}^{BTP} = -\dfrac{D_k^{DVA}}{d_k^{BTP}} - L_k^{BTP} \\ A_{k+1}^{iTraxx} = 0 \\ A_{k+1}^{SX7E} = 0 \end{cases}$$

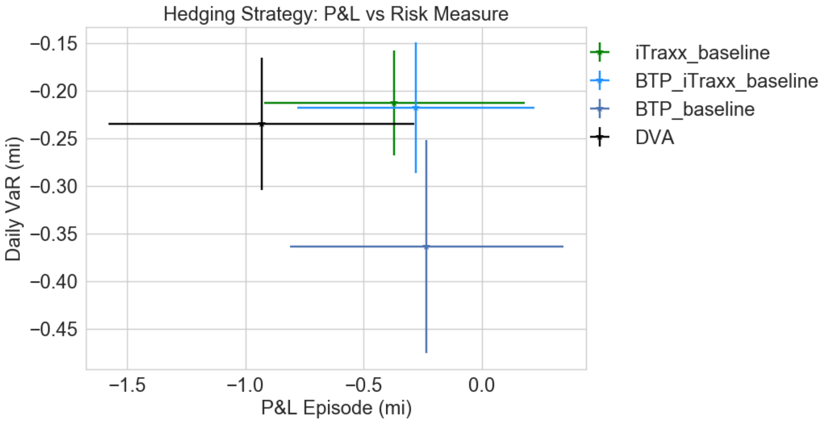
iTraxx Baseline

$$\begin{cases} A^{BTP} = 0 \\ A^{iTraxx} = -2\dfrac{D^{DVA}}{d^{iTraxx}} - L^{iTraxx} \\ A^{SX7E} = 0 \end{cases}$$

BTP-iTraxx Baseline

$$\begin{cases} A^{BTP} = -\dfrac{1}{2}\dfrac{D^{DVA}}{d^{BTP}} - L^{BTP} \\ A^{iTraxx} = -\dfrac{D^{DVA}}{d^{iTraxx}} - L^{iTraxx} \\ A^{SX7E} = 0 \end{cases}$$

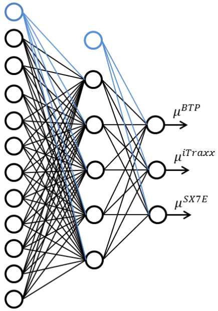
Baseline


Hedging Strategy: P&L vs Risk Measure
Daily VaR (mi)
−0.15
−0.20
−0.25
−0.30
−0.35
−0.40
−0.45
−1.5
−1.0
−0.5
0.0
P&L Episode (mi)
iTraxx_baseline
BTP_iTraxx_baseline
BTP_baseline
DVA

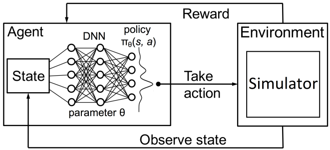
State - Action

$$s_i = [baseline\_features_i,\ total\_allocation_i,\ price_i]$$

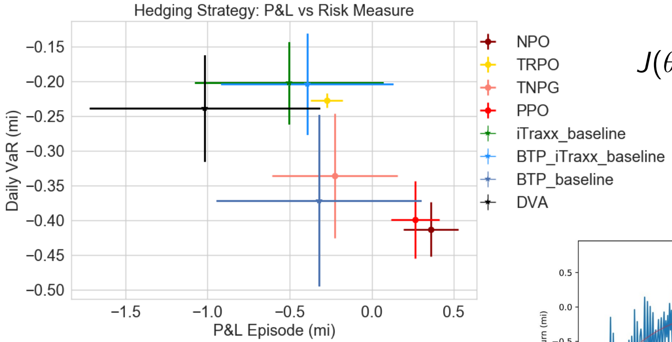
1. $baseline\_features_i$:
$$\left[\frac{D_i^{DVA}}{d_i^{BTP}},\ \frac{D_i^{DVA}}{d_i^{iTraxx}}\right]$$
2. $total\_allocation_i$:
$$\left[L_i^{SX7E},\ L_i^{BTP},\ L_i^{Bund},\ L_i^{ITRAXX},\ \psi_i^0\right]$$
3. $prices_i$:
$$\left[X_i^{SX7E},\ s_i^{BTP-Bund},\ X_i^{iTraxx},\ \pi_i^{5y}\right]$$



$$num\ param = 11 \cdot 5 + 5 \cdot 3 + 11 + 3 = 84$$

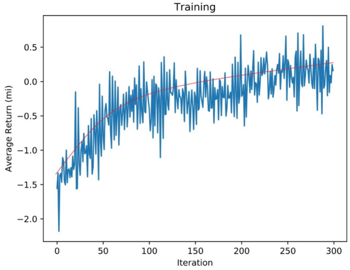


NPO, TRPO, PPO, TNPG - Train



$$J(\theta) = \mathbb{E}_{\pi_\theta}[\sum_{k=0}^{H} P\&L_k]$$

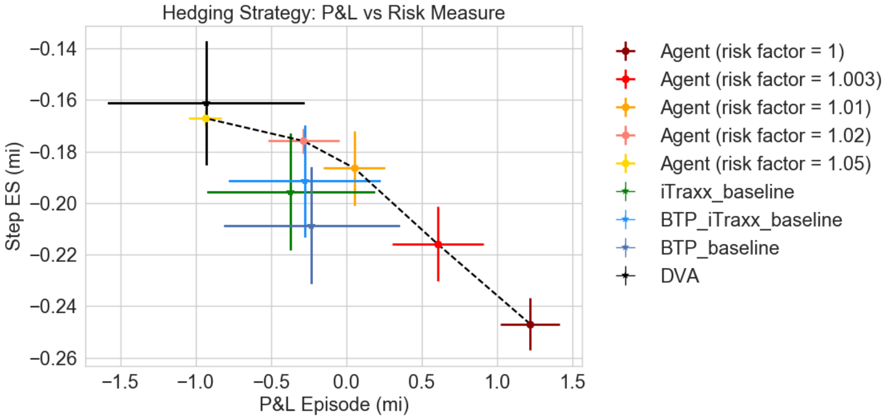
FFNN: one hidden layer, 5 neurons. Activation function: *tanh*.



NPO. Average return during training.

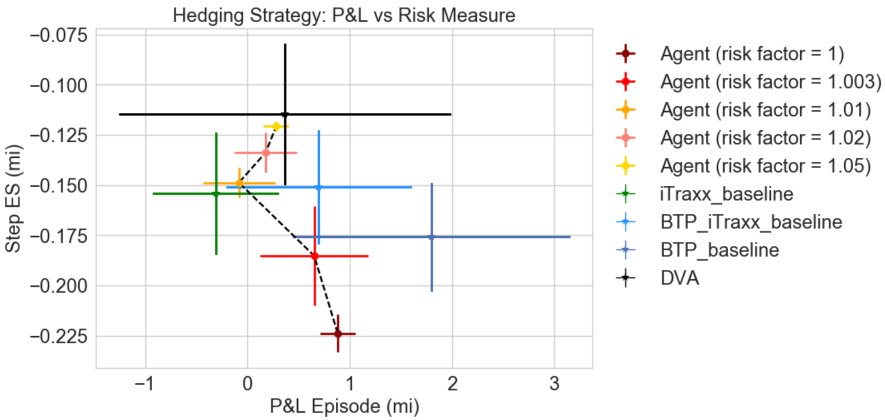
Efficient Frontier - NPO - Train

$$R(x) = \begin{cases} x & \text{if } x \geq 0 \\ 1 - (1-x)^{rf} & \text{if } x < 0 \end{cases} \quad \Rightarrow \quad J(\theta) = \mathbb{E}_{\pi_\theta}[\sum_{k=0}^{H} R(P\&L_k)]$$



Efficient Frontier - NPO - Test

$$R(x) = \begin{cases} x & \text{if } x \geq 0 \\ 1-(1-x)^{rf} & \text{if } x < 0 \end{cases} \quad \Rightarrow \quad J(\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{k=0}^{H} R(P\&L_k)\right]$$



Questions?

Email : giorgio.vit@jpmorgan.com
giorgio.vit123@gmail.com

Thesis : https://www.politesi.polimi.it/bitstream/10589/140092/3/2018_04_Vit.pdf