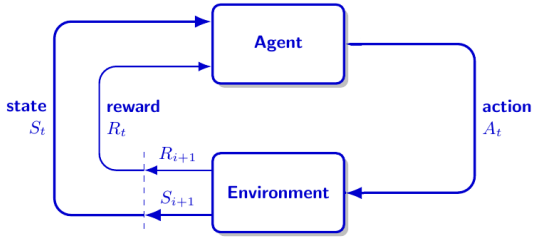
IMPERIAL



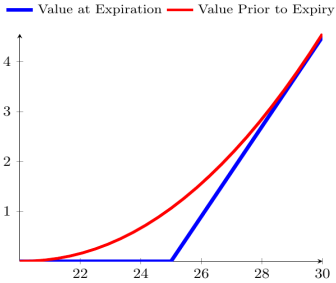
Deep Hedging

Under Capital Constraint

Guillaume Dechambre
CID: 02257544

Objectives

Two Deep Hedging Problems

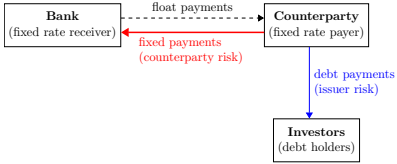


I. European Option

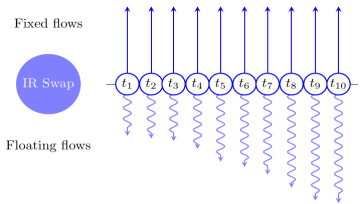
Hedging a European option provides a reference to compare other deep hedging publications with.

Objectives

- Design a novel reward which includes capital constraint and is insensitive to arbitrage.
- Tackle deep hedging on CVA of an IR swap.
- Benchmark our agent on ”real” trajectories.



Credit Valuation Adjustment quantifies the loss arising from counterparty risk on derivatives.



II. CVA on a new 5y IR Swap

An interest rate IR swap is a series of cashflows whereby a ”payer” pays the fixed rate at regular intervals and receives the floating rate from the ”receiver”.

The second task is for our agent to hedge the interest rate risk of the CVA of an IR swap.

European Option: Delta-Hedging

What Actions can the Agent Perform?

On the European option front, the stock price follows a Geometric Brownian Motion (GBM), with S_t as the stock price while its variance ν_t^S follows a CIR process following the Heston model [Hes93]:

Stock price: $dS_t = \mu^S S_t dt + \sqrt{\nu_t^S} S_t dW_t^S$

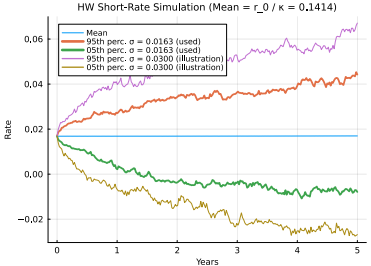
Variance: $d\nu_t^S = \kappa^\nu \left(\theta^\nu - \nu_t^S\right) dt + \xi^\nu \sqrt{\nu_t^S} dW_t^\nu$ with $\rho_{S,\nu^S} = -0.5$, see [MWB+22]

$$\underbrace{\frac{\partial C}{\partial t}}_{\text{Change from time}} + \underbrace{\frac{1}{2}\frac{\partial^2 C}{\partial S_t^2}\sigma^2 S_t^2}_{\text{Change from 2nd derivatives}} - \underbrace{rC_t}_{\text{Grows at risk free rate}} + \underbrace{r\frac{\partial C}{\partial S_t}S_t}_{\text{Change from stock price move}} = 0 \text{ subject to : } C_T = \max(S_T - K)$$

Delta-Hedging strategy (D-H) involves neutralizing, at each step, the first order sensitivity or delta. This serves as a benchmark for our agent's performance. For the European option problem, the agent can decide to buy or sell shares at each trading opportunity which happens 6 times a day.

CVA Building Blocks

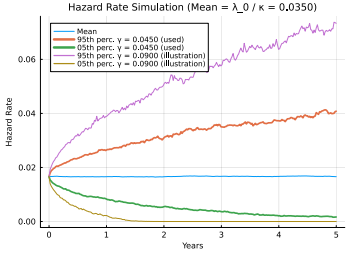
Three Models to Simulate CVA



Interest Rate Simulation

An interest rate model is necessary to generate trajectories.

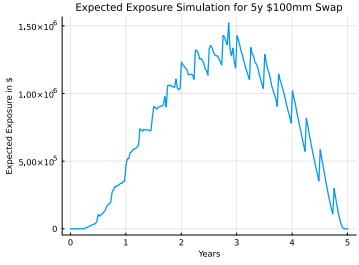
2 models proposed: **I** one-factor allowing negative rates or **II** 4-factors producing positive rates.



Credit Model

The credit simulation model produces a hazard rate which is equivalent to the probability of default of the counterparty.

We choose a model which does not allow negative hazard rates.



Exposure

Positive expected exposure consists of the expected amount a counterparty owes a bank in the future.

Hull & White model (one-factor) was chosen for Monte Carlo (MC) exposure generation.

CVA: Delta-Hedging

Model I: named "normal" [N] model.

Interest rate: $\mathrm{d}r_t = \kappa^r \left(\theta_t^r - r_t\right) \mathrm{d}t + \sigma^r \mathrm{d}W_r^{\mathbb{Q}}$, the H-W short rate [HW90]

Hazard rate: $\mathrm{d}\lambda_t = \kappa^\lambda \left(\theta^\lambda - \lambda_t\right) \mathrm{d}t + \gamma^\lambda \sqrt{\lambda_t} \mathrm{d}W_t^\lambda$, the CIR model [CIR85]

Model II: named "log-normal" [LN] model.

Interest rate: $\forall j \in \{1, 3, 5, 7\}$ the swap rate S_t has the dynamic $\frac{\mathrm{d}S_t^j}{S_t^j} = \sigma^j \mathrm{d}W_t^j$ akin to [Bla76]

Hazard rate: $\frac{\mathrm{d}\lambda_t}{\lambda_t} = \sigma^\lambda \mathrm{d}W_t^\lambda$

$$\mathrm{dCVA} = \underbrace{\frac{\partial \mathrm{CVA}}{\partial t}\mathrm{d}t}_{\text{change with time}} + \underbrace{\sum_{i \in \{1y,3y,5y,7y\}} \frac{\partial \mathrm{CVA}}{\partial s_i}\mathrm{d}s_i}_{\text{change from swap rates } s_i} + \underbrace{\frac{\partial \mathrm{CVA}}{\partial \lambda^{\mathrm{Hzd}}}\mathrm{d}\lambda^{\mathrm{Hzd}}}_{\text{change from hazard rate (assumed hedged)}} + \underbrace{\cdots}_{\text{higher order changes \& cross effects}}$$

Delta-Hedging strategy (D-H): Each day this strategy involves neutralizing the first order IR sensitivities to swap rates. There are 4 swap tenors $\{1y, 3y, 5y, 7y\}$ available to the agent to trade at each step.

Designing The Reward

Capital + Risk Aversion + Transaction Cost

Capital: $K_t^{IR} = \sqrt{(\delta_{IR}^{P} - \delta_{IR}^{H})^{T} C_{IR} (\delta_{IR}^{P} - \delta_{IR}^{H}) + R \left(\sum_{i \in \{1y,3y,5y,7y\}} \delta_{IR,i}^{H} \right)^2}$ and $\frac{dA_t^K}{dt} = r_K \times K_t$, capital acct. A_t^K.

Risk aversion: $\mathbb{SD}[Z] := \left(\mathbb{E}\left[(Z - \mathbb{E}[Z])_-^2 \right] \right)^{\frac{1}{2}}$, the semi-deviation of a random variable Z.

Transaction cost: $c(s_t, a_t) = \alpha |a_t| |h_t|$ where h_t where h_t is the sensitivity of a standard hedge, a_t the action taken and α the transaction cost per unit.

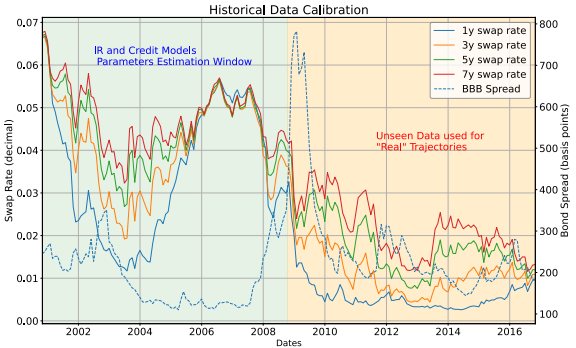
The semi-deviation is “an appealing alternative to the standard deviation which does not distinguish between the variability of upside and downside variations.” [TCGM17] Semi-deviation is a practical alternative to cVar (expected shortfall), which has the property of being a coherent risk measure.

Put together, the agent’s reward per step takes the form:

$$R_t = r_K K_t dt + c(s_t, a_t) + \beta \times \mathbb{SD}[MTM_{[t-L,t]}]$$

where $MTM_{[t-L,t]}$ is the MTM over the past L steps, L is the look-back period and the parameter β projects the risk aversion onto the rest of the reward components, setting its relative weight in the process.

Historical Calibration For The CVA Problem

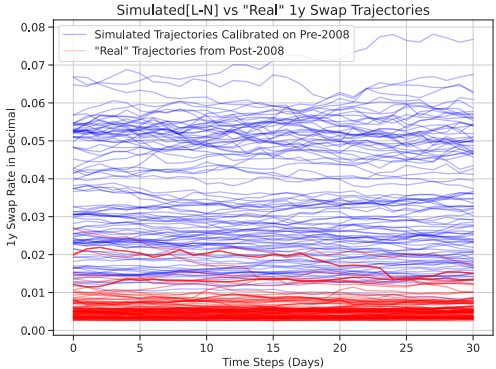


Historical simulation parameters estimation or ”calibration” covers daily US swap rates and BBB-rated spreads (source FRED [oGotFRSU24]) over the period 2000-2008.

We also sample ”real” 30 day long trajectories every 10 days over the 2008-2016 period.

Below one can see the simulated log-normal [LN] 30 days paths against the ”real” trajectories sampled from the 2008-2016 period.

There is very little overlap between the two, making the ”real” test that much harder for the agent.



Generating Trajectories Offline

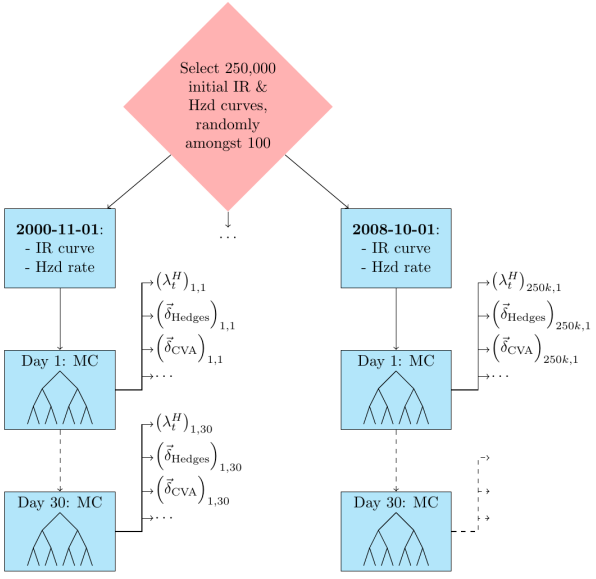
Monte Carlo Square Framework

We generate 250,000 trajectories for each CVA environment using randomly selected curves (with replacement) taken from the beginning of each month from 2000 to 2008.

Each of these trajectories are comprised of 30 days for training, and 90 days for testing.

For each of these steps, we calculate CVA on an interest swap, its sensitivities to swap rates (4 simulations) and its sensitivity to hazard rate which makes the whole calculation Monte Carlo square.

We used the Julia language to generate this simulation where we saw about 50x speed-up compared to the same implementation in Python.



Simulating trajectories for training and simulated testing.

Policy Gradient Algorithm

Algorithm 1 PPO-Clip: Proximal Policy Optimization in [SWD+17], [Ach18]

1: **Initialization:** Initial policy parameters θ_0, initial value function parameters ϕ_0
2: **for** $k \in \{1, ..., n\}$ **do**
3: Collect set of trajectories $\mathcal{D}_k = \{\tau_i\}$ by running policy $\pi_k = \pi(\theta_t)$ in the environment
4: Compute rewards per step $\hat{R}_t$.
5: Compute advantage estimates, $\hat{A}_t$ using GAE based on the current value function V_{ϕ_k}.
6: Update the policy by maximizing the PPO-Clip objective:
Update the policy by maximizing the PPO-Clip objective (via stochastic gradient ascent with Adam):

$$\theta_{k+1} = \arg\max_{\theta} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \min \left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} A^{\pi_{\theta_k}}(s_t, a_t), g\left(\epsilon, A^{\pi_{\theta_k}}(s_t, a_t)\right) \right)$$

with $g(\epsilon, A) = (1+\epsilon)\, A$ if $A \geq 0$ and $g(\epsilon, A) = (1-\epsilon)\, A$ if $A < 0$

7: Fit value function by regression on mean-squared error:

$$\phi_{k+1} = \arg\min_{\phi} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^{T} \left(V_\phi(s_t) - \hat{R}_t \right)^2$$

8: Using a gradient descent algorithm.
9: **end for**

Neural Networks Architecture

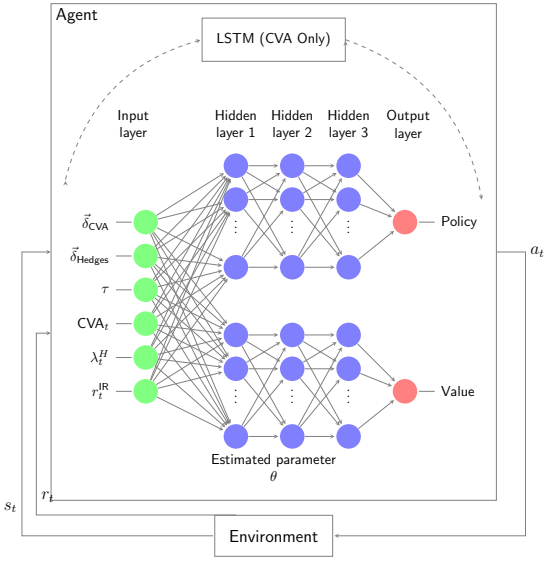
Policy and Value Networks

The **policy** (actor) and **value** (critic) are both modelled by neural networks. They share the same inputs.

European option: Each network is represented by a 3-layers 12-units multilayer perceptron (MLP).

CVA: In the case of CVA, we found that a long short-term memory (LSTM, [HS97]) architecture with 2 layers and 128 units produced substantially better results than its vanilla feed-forward counterpart.

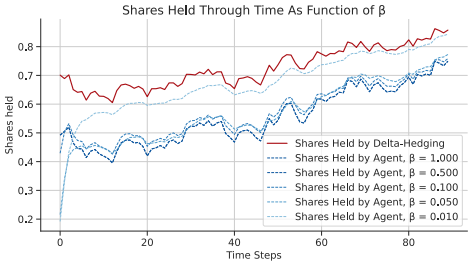
Activation function is hyperbolic tangent in both cases.



Summarized policy and value networks

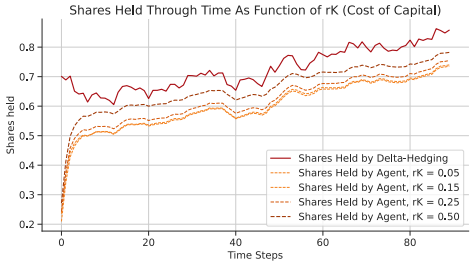
European Option Results

Zooming-In on a Sim 90 Days Trajectory



Shares Held as Function of β (Risk Aversion)
When β increases, the agent's actions mirror the delta-hedging but at a lower level of holding.

This behaviour is due to the negative correlation ρ_{S,ν^S} = -0.50 between the volatility process and the stock price.



Shares Held as Function of r_K (Cost of Capital)
Delta-hedging strategy has zero capital footprint given the capital formula.

Therefore, as r_K increases, the agent's strategy converges to the delta-hedging strategy since the latter is optimal at reducing capital.

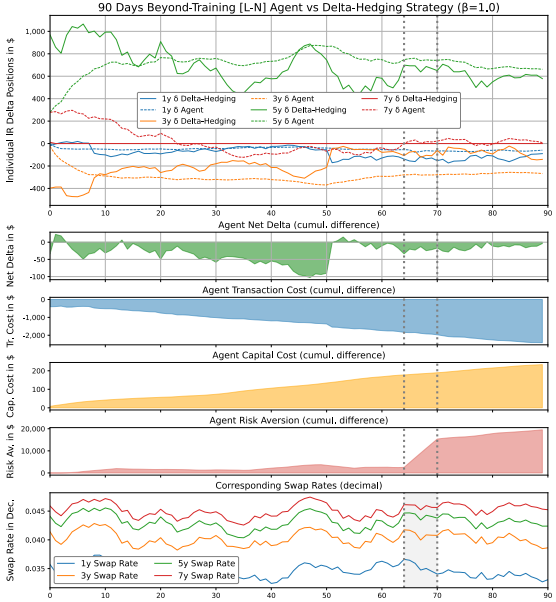
CVA Results

Zooming-In on a Sim 90 Days Trajectory

The [LN] agent's actions are smoother than D-H, which allows it to save on transaction costs compared to D-H, while under-performing it on risk aversion and slightly on the capital front (see the orange area chart).

The agent keeps some risks open at the individual delta level, but overall stays well hedged (see the green area chart).

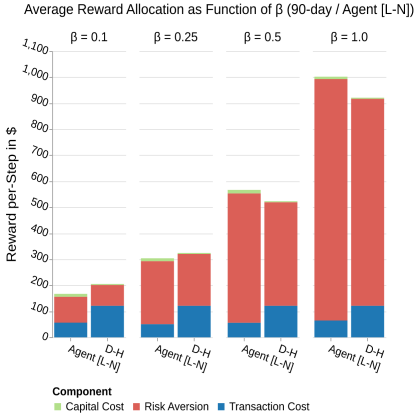
At day 64, a sudden downward move of the 1y swap rate makes creates a P&L event which leads the agent to under-perform D-H on the risk aversion (see the red area chart).



Specific simulated 90 days test trajectory for the CVA problem

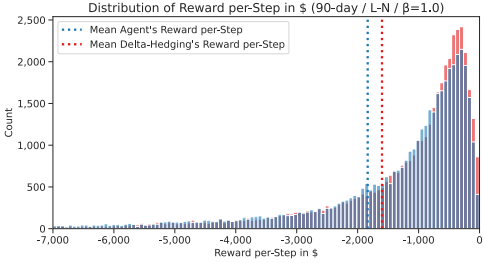
CVA Results

Overall Test Results on Sim 90 Days Trajectories



Cost Allocation

As we suspect, the agent gains on transaction cost, but loses on capital and risk aversion.



Reward Distribution

Red bars indicate the rewards of D-H and blue bars the performance of the [LN] agent.

Despite being trained on 30 days horizon, the log-normal [LN] agent's per-step reward distribution overlaps the delta-hedging's counterpart on 90 days tests while slightly under-performing it for $\beta = 1.00$.

CVA Results

Overall Test Results on 200 "Real" 30 Days Trajectories

30-days Real dataset	Agt. [LN]	Agt. [LN+LSTM]	Agt. [N+LSTM]	Delta-Hedging
β=0.10, r_K=0.05	-666	**-211**	-412	-280
β=0.10, r_K=0.15	-698	**-227**	-306	-285
β=0.10, r_K=0.25	-629	**-237**	-287	-290
β=0.10, r_K=0.50	-852	**-253**	-273	-302
β=0.10, r_K=0.15	-698	**-227**	-306	-285
β=0.25, r_K=0.15	-997	**-406**	-616	-421
β=0.50, r_K=0.15	-1,756	-727	-1,239	**-648**
β=1.00, r_K=0.15	-3,398	-1,215	-2,707	**-1,102**

The log-normal [LN] agent outperforms the normal [N] agent in all scenarios on the ”real” dataset. This may be due to the one factor nature of the latter. The [LN] agent does better than delta-hedging when β is not too large, e.g. less than 0.50.

The LSTM architecture makes a significant positive contribution to the agent’s performance.

When β becomes large, delta-hedging dominates as there is no stochastic volatility or market-credit correlation the agent can learn to gain an advantage.

IMPERIAL

Thank you.
Any questions?

References I

[Ach18] Joshua Achiam, Spinning Up in Deep Reinforcement Learning.

[Bla76] Fischer Black, The pricing of commodity contracts, Journal of Financial Economics **3** (1976), no. 1, 167–179.

[CIR85] John C Cox, Jr Ingersoll, Jonathan E, and Stephen A Ross, A Theory of the Term Structure of Interest Rates, Econometrica **53** (1985), no. 2, 385–407.

[Hes93] Steven L. Heston, A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, The Review of Financial Studies **6** (1993), no. 2, 327–343.

[HS97] Sepp Hochreiter and Jürgen Schmidhuber, Long short-term memory, Neural Comput. **9** (1997), no. 8, 1735–1780.

[HW90] John Hull and Alan White, Pricing Interest-Rate-Derivative Securities, The Review of Financial Studies **3** (1990), no. 4, 573–592.

[MWB+22] Phillip Murray, Ben Wood, Hans Buehler, Magnus Wiese, and Mikko S. Pakkanen, Deep hedging: Continuous reinforcement learning for hedging of general portfolios across multiple risk aversions, 2022.

[oGotFRSU24] Board of Governors of the Federal Reserve System (US), Xx-year swap rate (discontinued) [dswpxx], 2024, retrieved from FRED, Federal Reserve Bank of St. Louis.

[SWD+17] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov, Proximal policy optimization algorithms, CoRR **abs/1707.06347** (2017).

[TCGM17] Aviv Tamar, Yinlam Chow, Mohammad Ghavamzadeh, and Shie Mannor, Sequential decision making with coherent risk, IEEE Transactions on Automatic Control **62** (2017), no. 7, 3323–3338.