Algorithmic Trading with Reinforcement Learning

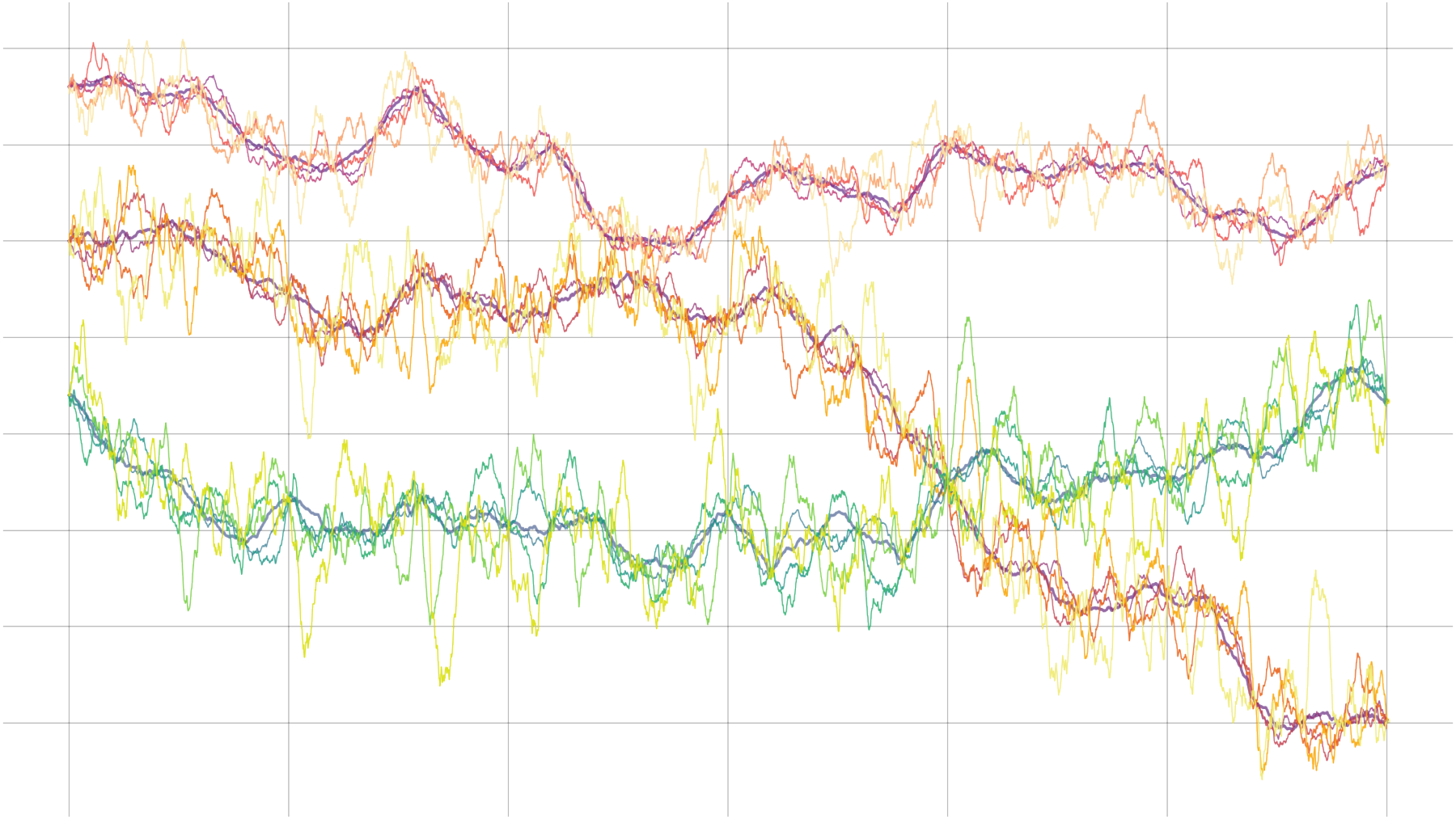
Second semester report

Leonardo Toffalini

2026-06-03

Problem statement

Fractional Brownian motion



Market model

Market model [1]:

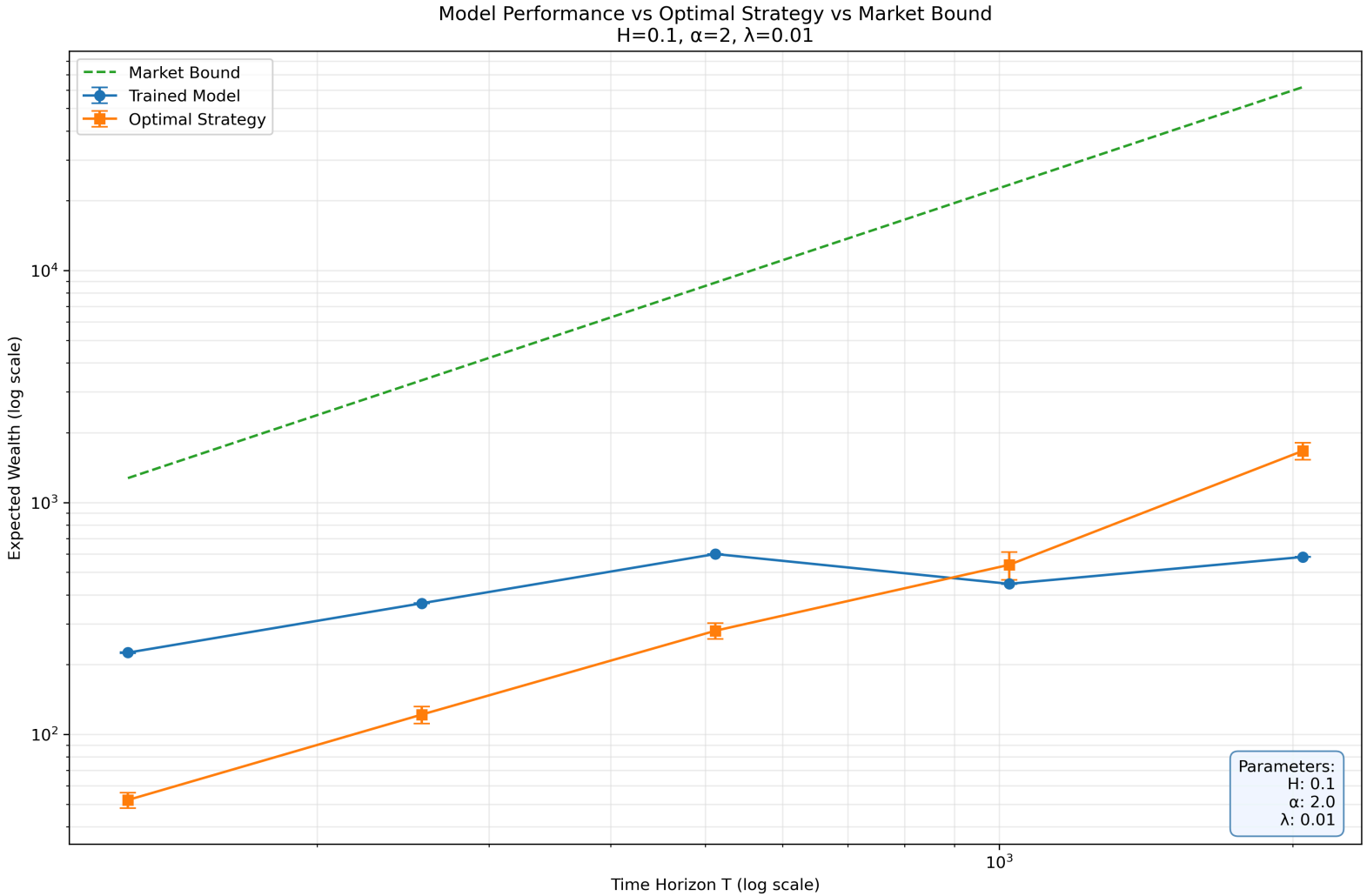
$$X_t^1(\phi) := z^1 + \int_0^t \phi_u \, \mathrm{d}u \qquad \text{(risky)}$$

$$X_t^0(\phi) := z^0 - \int_0^t \phi_u S_u \, \mathrm{d}u - \int_0^t \lambda |\phi_u|^\alpha \, \mathrm{d}u \qquad \text{(riskless)}$$

Goal:

$$\max_{\phi \in S(t)} \mathbb{E}\left[X_T^0(\phi)\right]$$

Previous work

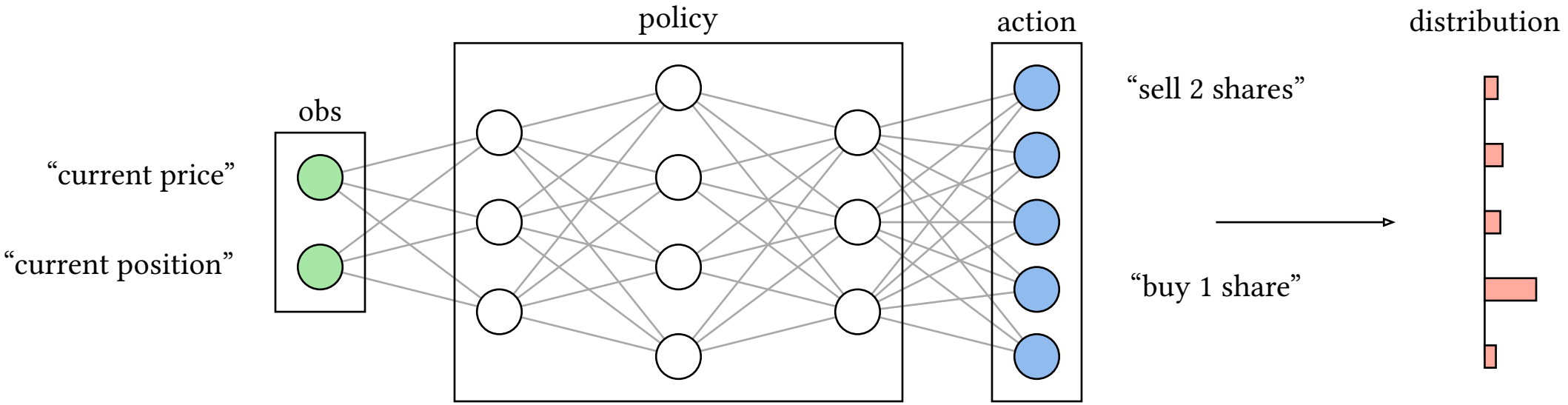

Model Performance vs Optimal Strategy vs Market Bound
H=0.1, α=2, λ=0.01
Market Bound
Trained Model
Optimal Strategy
Expected Wealth (log scale)
10^4
10^3
10^2
10^3
Time Horizon T (log scale)
Parameters:
H: 0.1
α: 2.0
λ: 0.01

1st semester

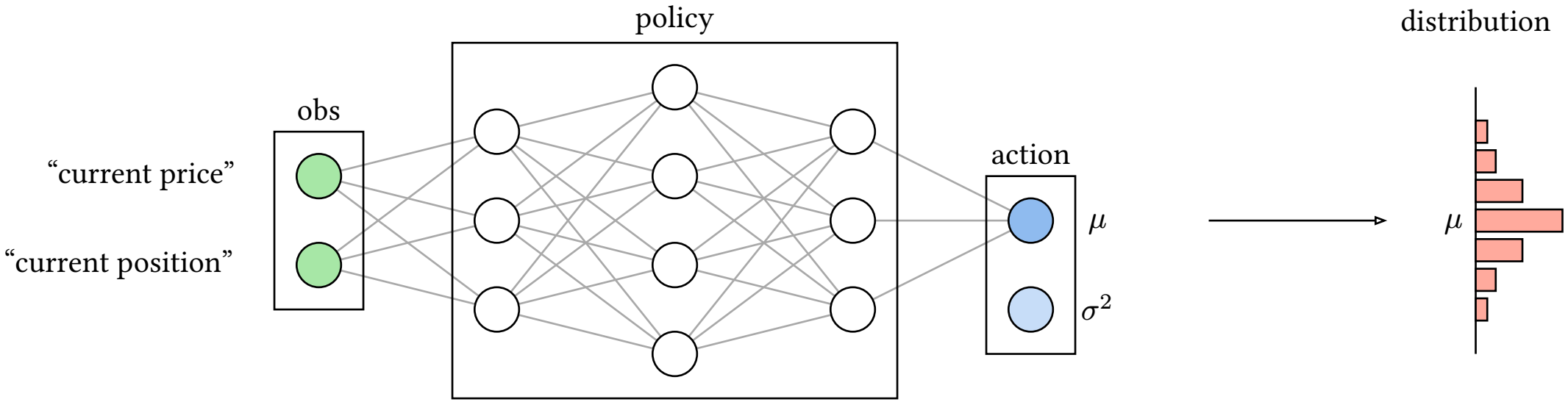
- The code base was rewritten in C from Python [2]
- Liquidation strategy is now a linear schedule
- Smarter rewards: anticipating liquidation value
- Three orders of magnitude speedup (1.5k $\Rightarrow$ 1.5M SPS)
- Large-scale hyperparameter search with CARBS [3]
- Better performance for fixed time horizons

Current work

Discrete action policy schematic diagram



Continuous action policy schematic diagram



Accurate continuous action policy diagram

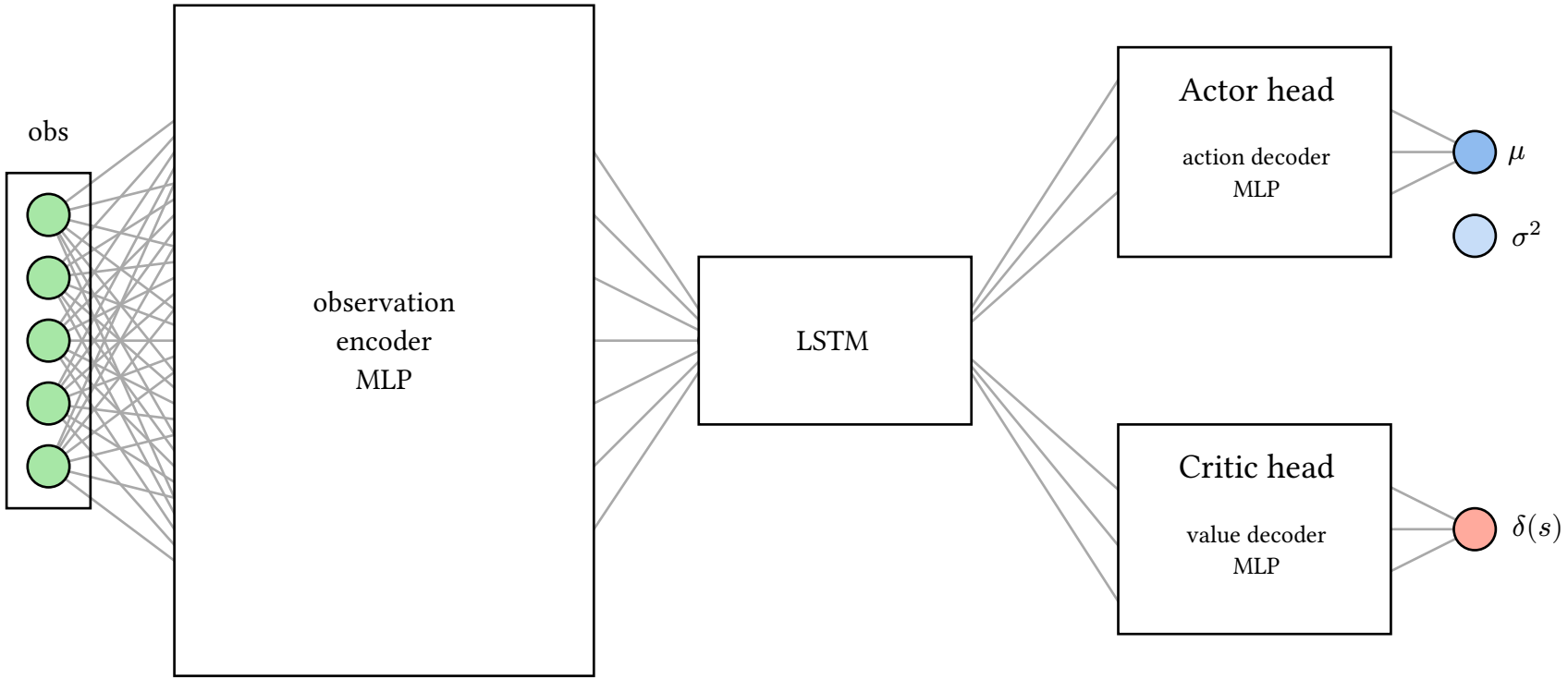


Figure 5: Policy network architecture consisting of an observation encoder, recurrent LSTM core, and actor-critic output heads [4].

Analytic strategy vs market bound

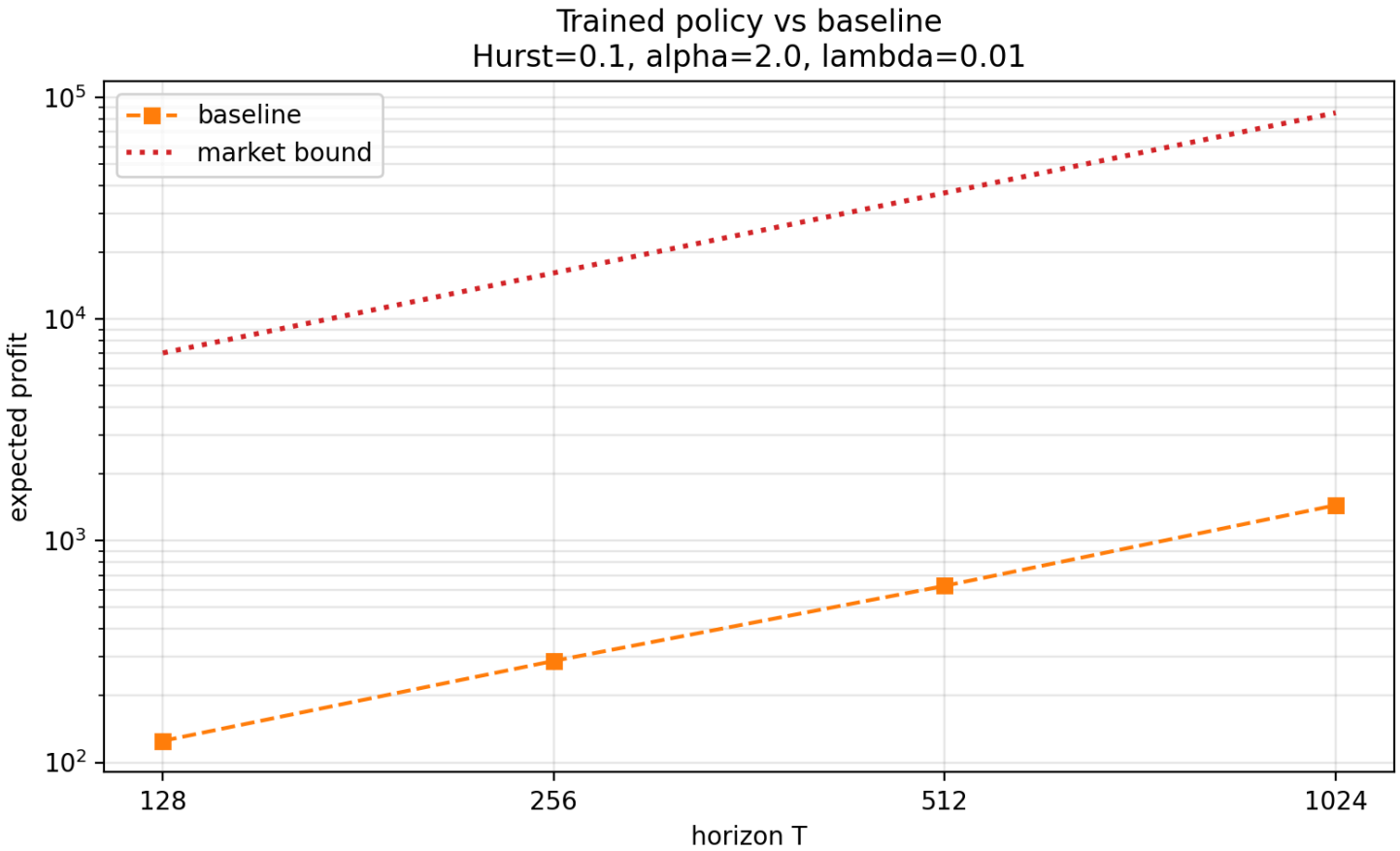


Figure 6: Learnt policy versus asymptotically optimal strategy versus market bound [1]. Each dot represents the mean return of 500 rollouts.

Analytic strategy vs trained policy

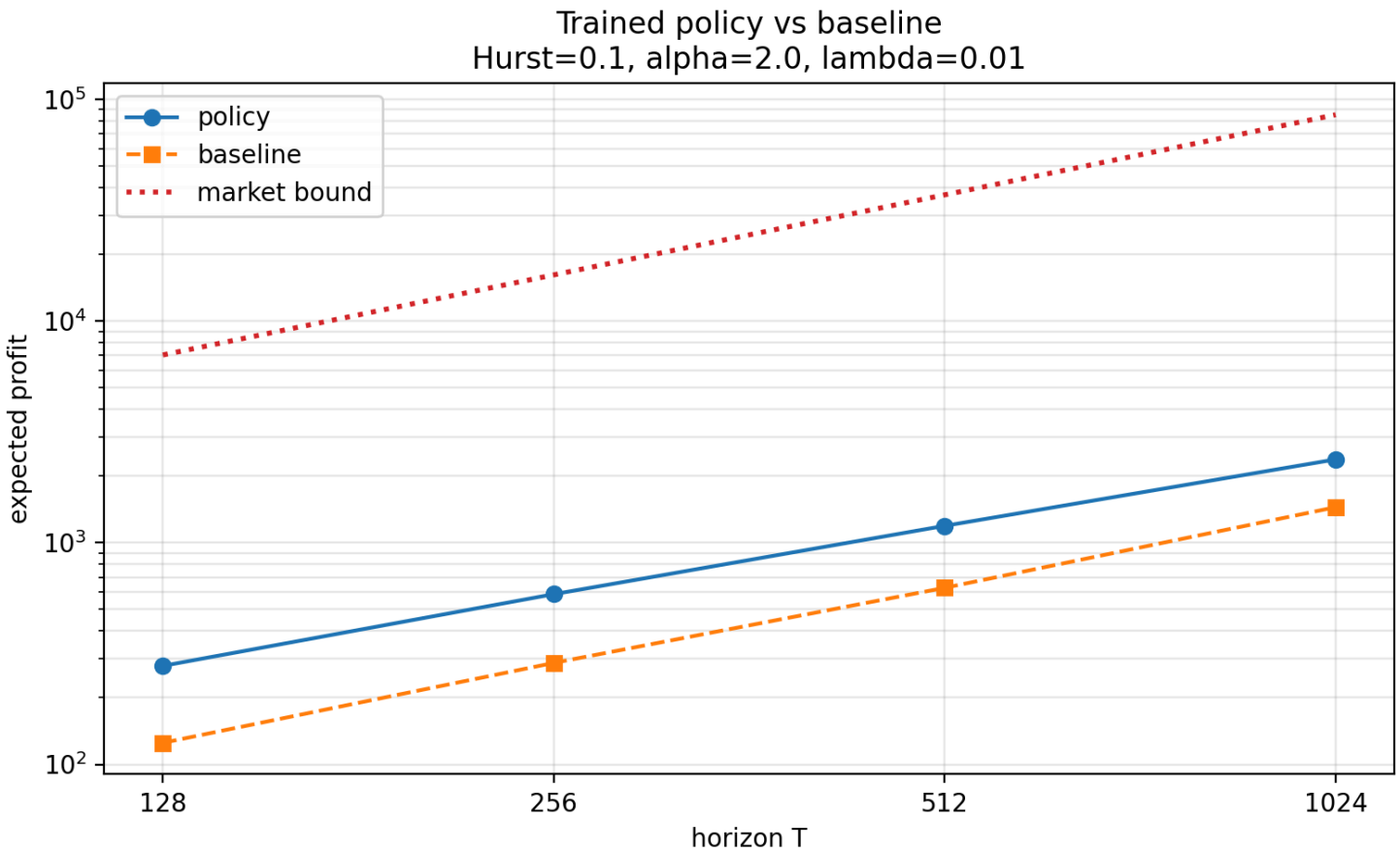


Figure 7: Learnt policy versus asymptotically optimal strategy versus market bound. Each dot represents the mean return of 500 rollouts.

Analytic strategy vs trained policy

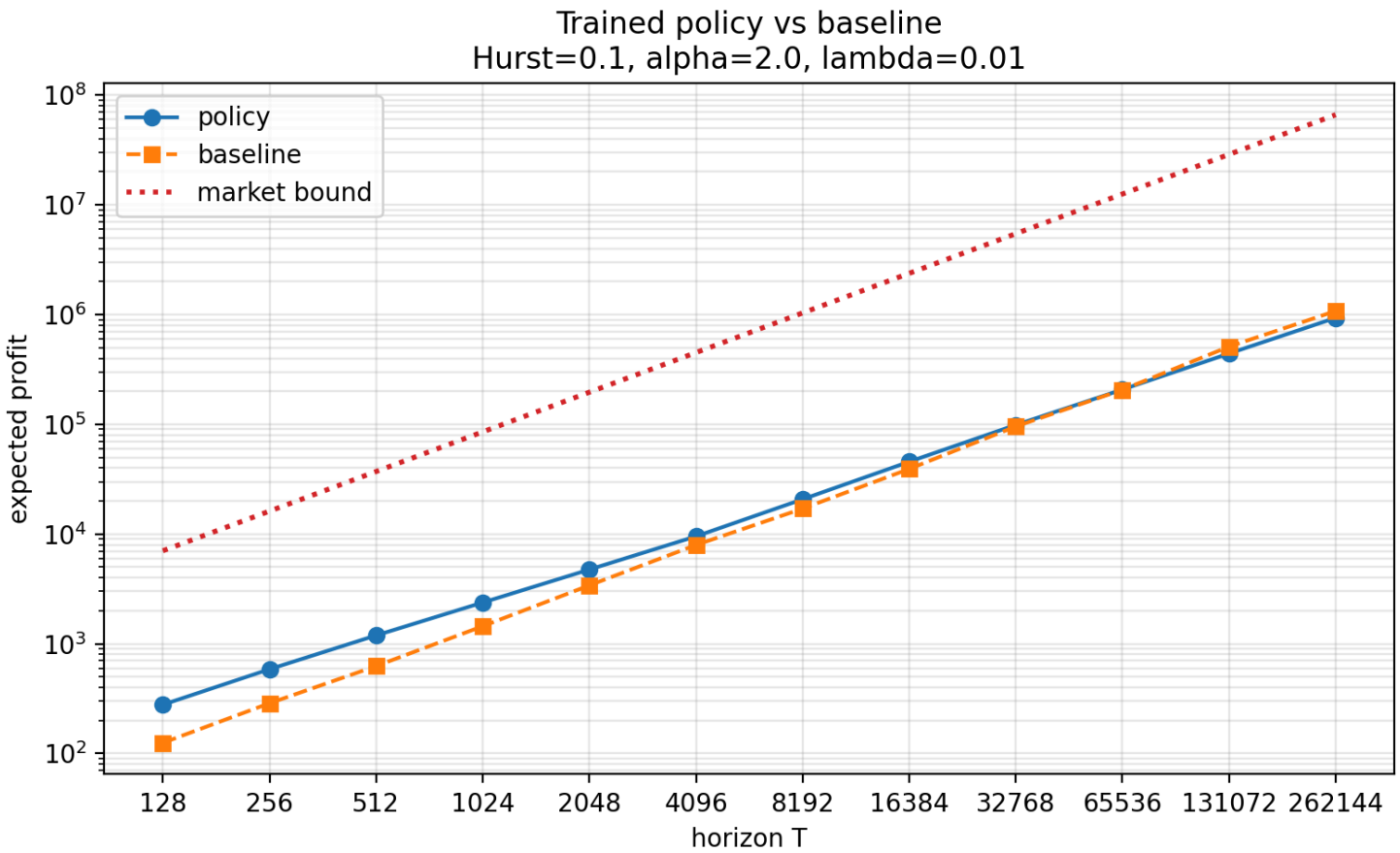


Figure 8: Learnt policy versus asymptotically optimal strategy versus market bound. Each dot represents the mean return of 500 rollouts.

Analytic strategy vs trained policy

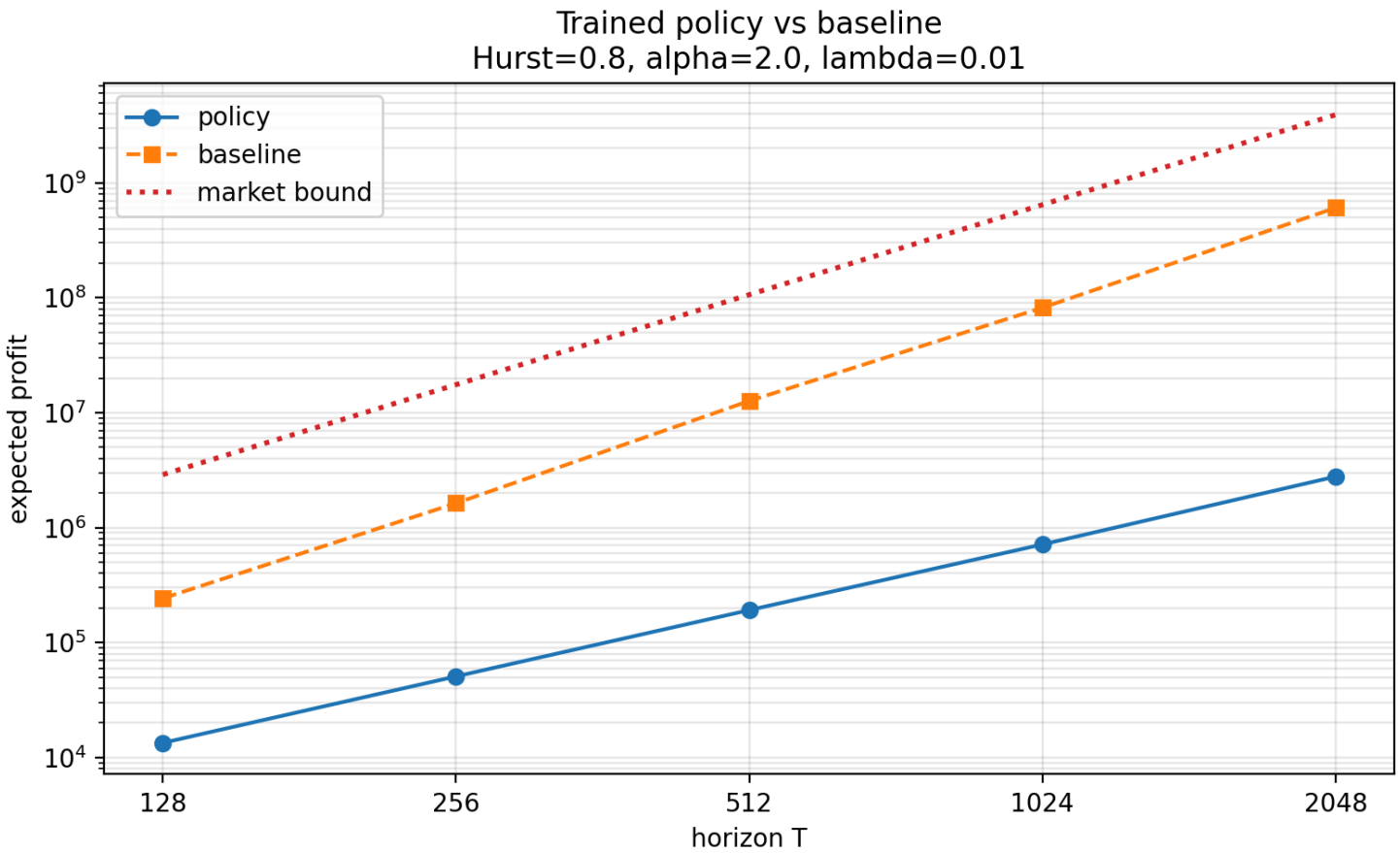


Figure 9: Learnt policy versus asymptotically optimal strategy versus market bound. Each dot represents the mean return of 500 rollouts.

Analytic strategy vs trained policy

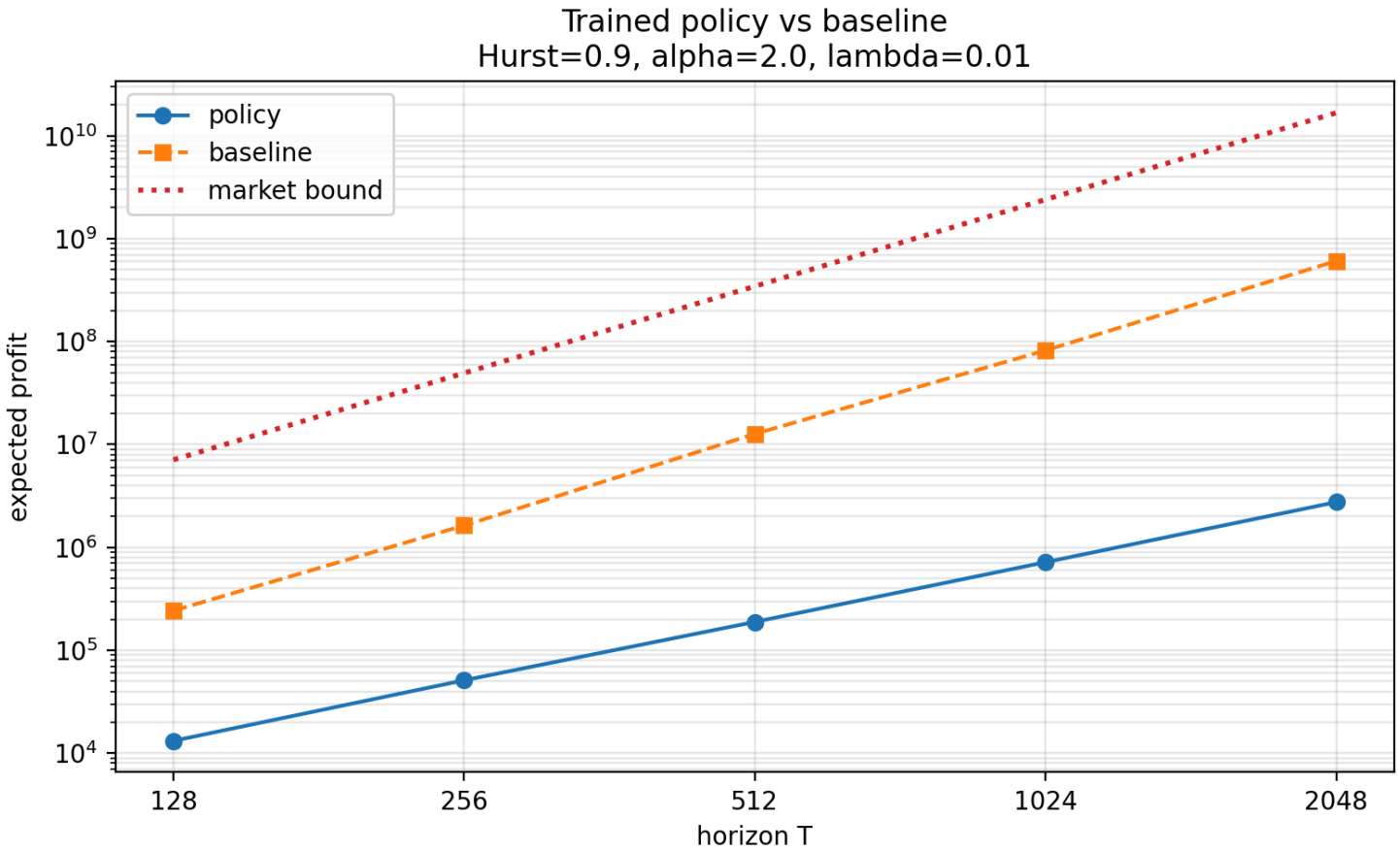


Figure 10: Learnt policy versus asymptotically optimal strategy versus market bound. Each dot represents the mean return of 500 rollouts.

Example rollouts

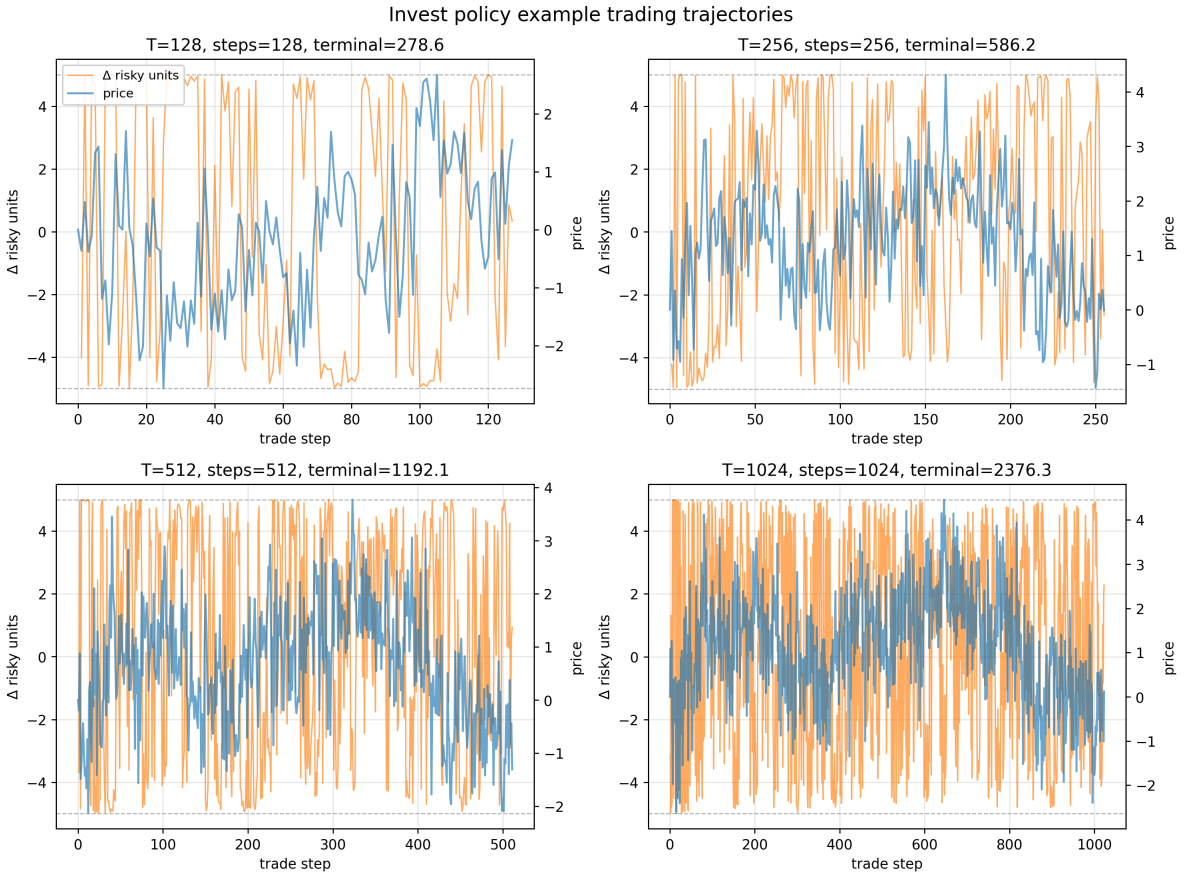


Figure 11: Example rollouts showing only the action – delta riskless units – and the fBm prices for increasing time horizons. The achieved return is noted in the subtitle of each subplot.

Example rollouts

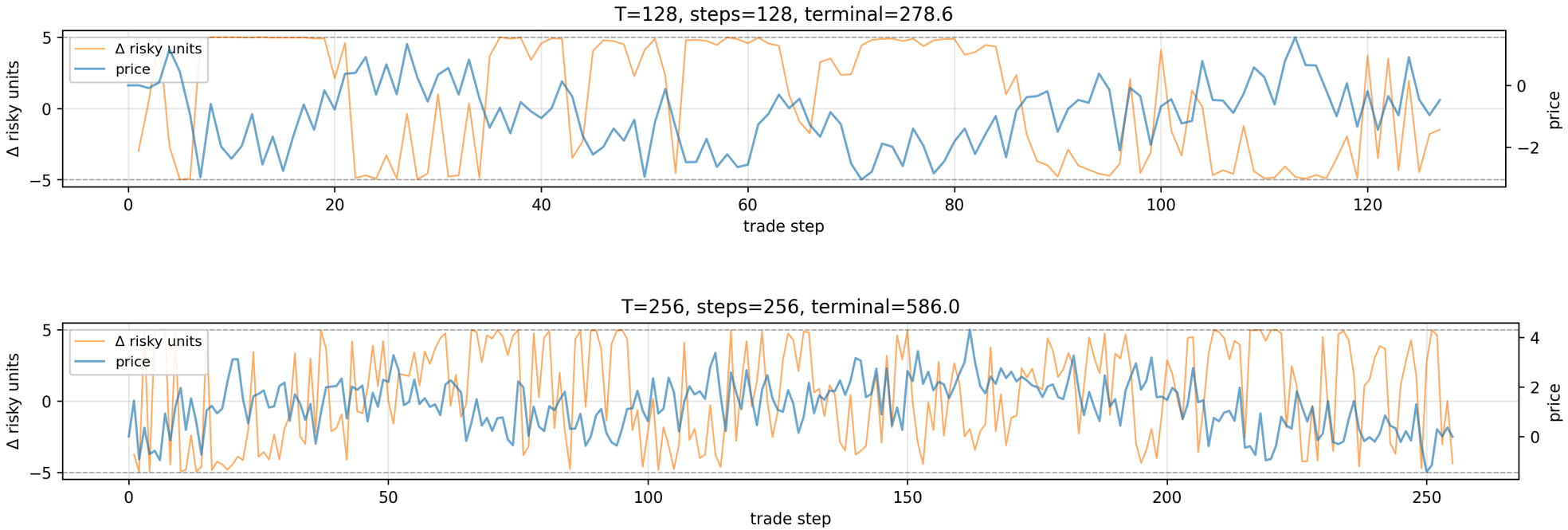


Figure 12: Example rollouts zoomed in showing only the action and the fBm prices.

Autocorrelation of the strategy

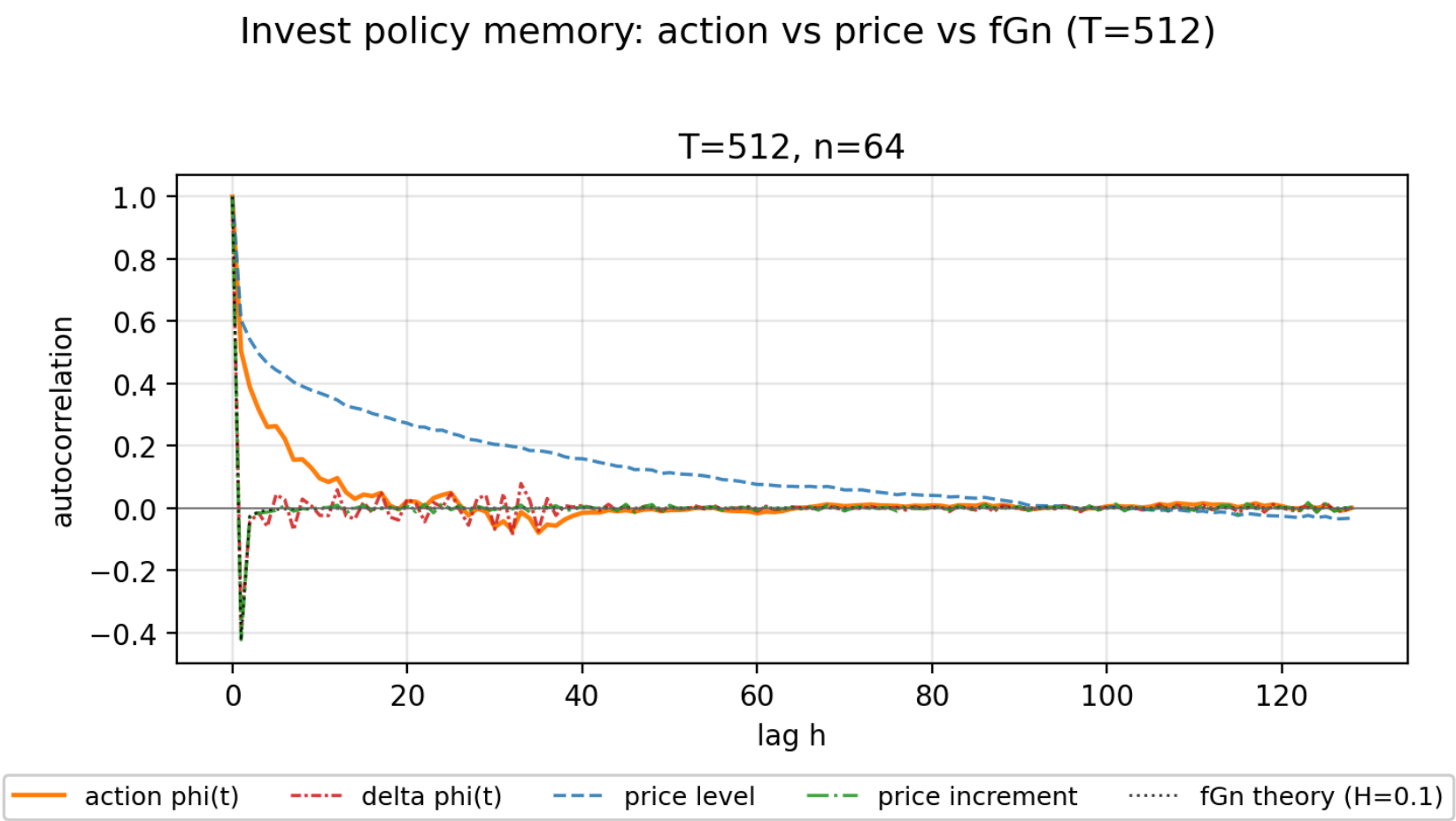


Figure 13: Autocorrelation of the strategy $\phi(t)$

Autocorrelation of the strategy

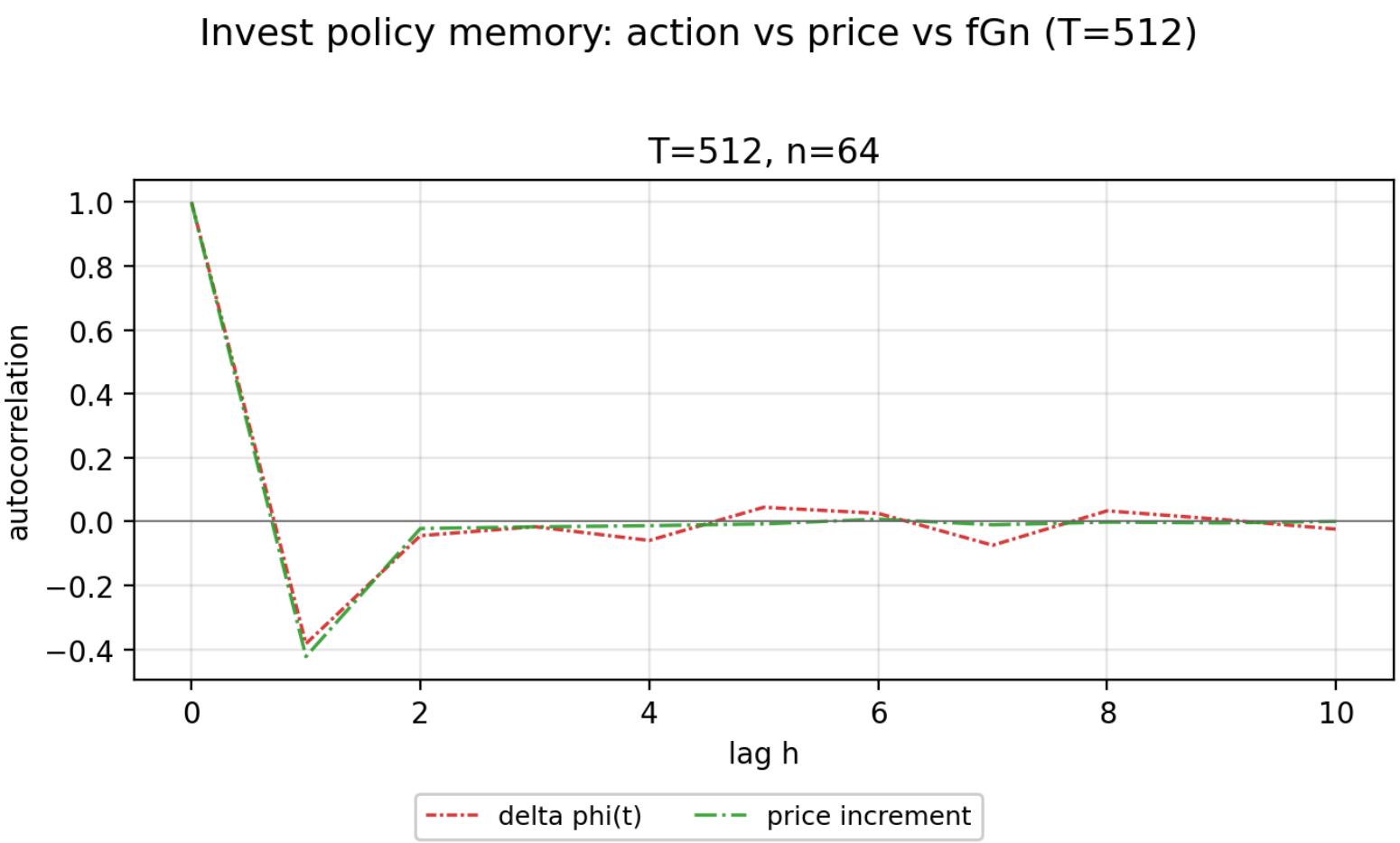


Figure 14: Autocorrelation of the strategy $\phi(t)$

Skewed time horizon distribution

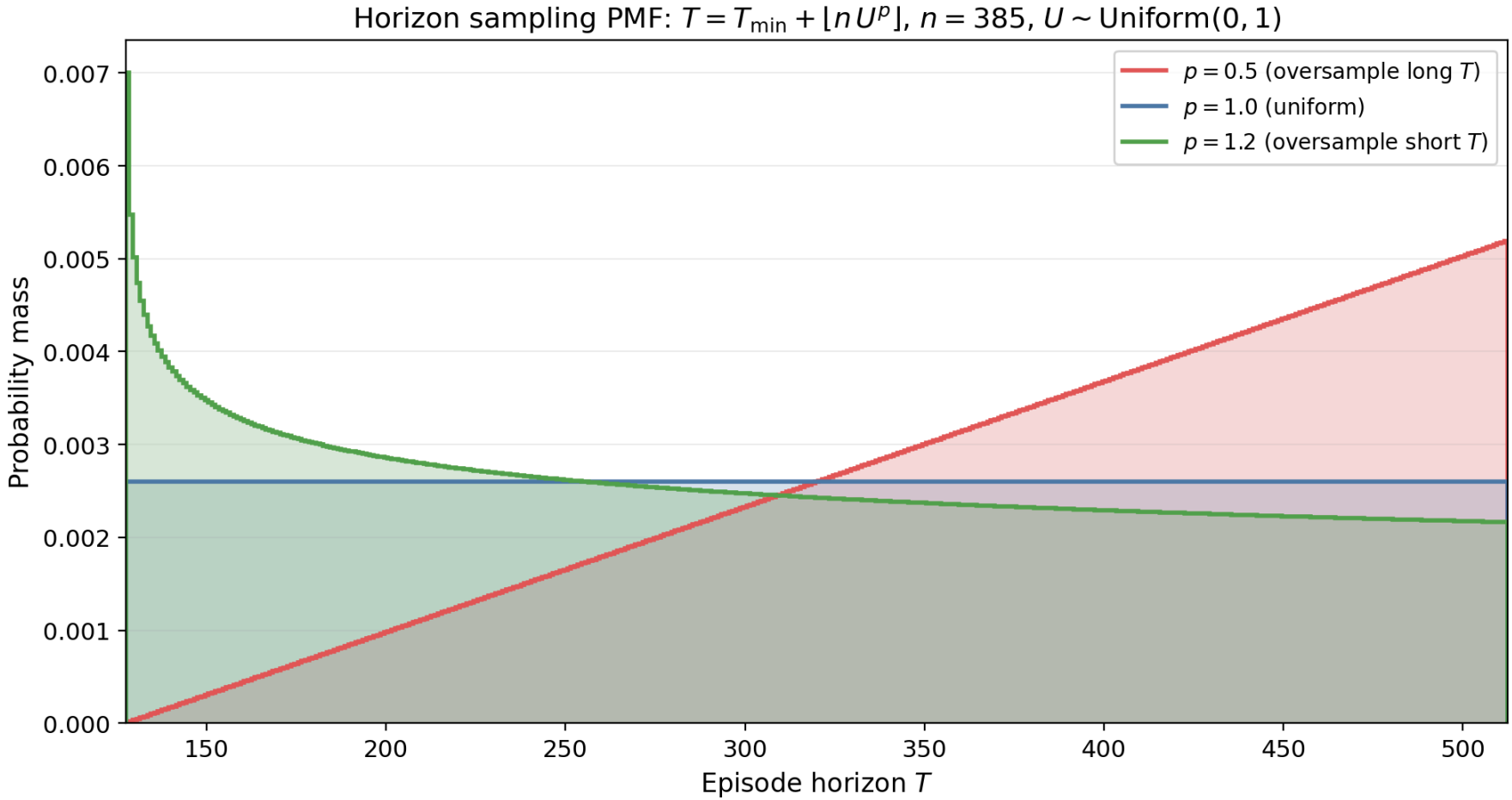


Figure 15: Distribution of time horizons with respect to p [5].

Skewed time horizon distribution

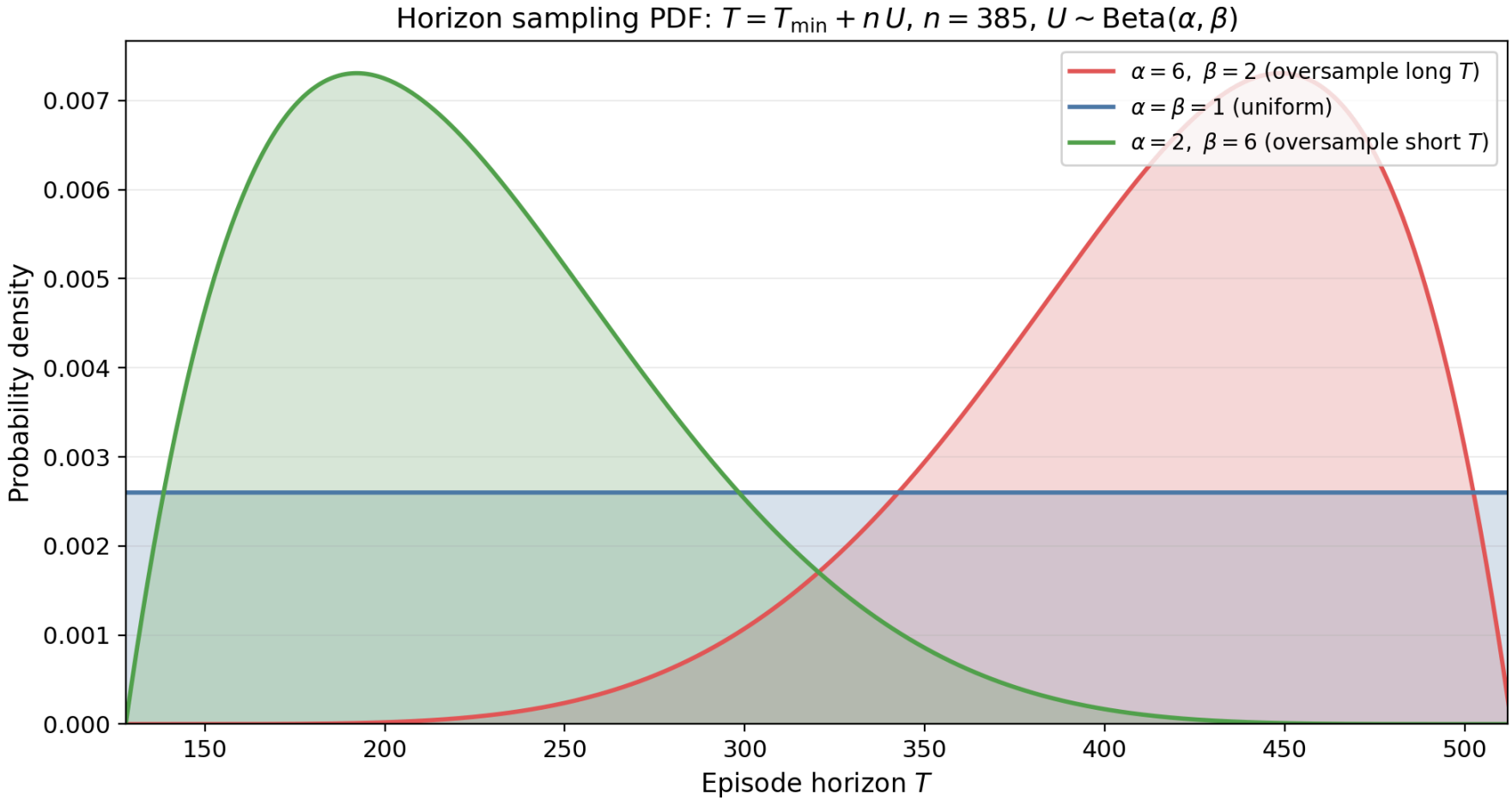


Figure 16: Distribution of time horizons with respect to α and β [5].

Bibliography

[1] P. Guasoni, Z. Nika, and M. Rásonyi, "Trading fractional Brownian motion," *SIAM journal on financial mathematics*, vol. 10, no. 3, pp. 769–789, 2019.

[2] J. Suarez, "PufferLib 2.0: Reinforcement Learning at 1M steps/s," *Reinforcement Learning Journal*, vol. 6, pp. 1378–1388, 2025.

[3] A. J. Fetterman *et al.*, "Tune As You Scale: Hyperparameter Optimization For Compute Efficient Training," *arXiv e-prints*, 2023.

[4] C. Berner *et al.*, "Dota 2 with large scale deep reinforcement learning," *arXiv preprint arXiv:1912.06680*, 2019.

[5] A. Elsafi, "Evaluating Domain Randomization Techniques in DRL Agents: A Comparative Study of Normal, Randomized, and Non-Randomized Resets.," *Computer Modeling in Engineering & Sciences (CMES)*, vol. 144, no. 2, 2025.

[6] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[7] S. Huang *et al.*, "CleanRL: High-quality Single-file Implementations of Deep Reinforcement Learning Algorithms," *Journal of Machine Learning Research*, vol. 23, no. 274, pp. 1–18, 2022, [Online]. Available: http://jmlr.org/papers/v23/21-1342.html

[8] R. S. Sutton, A. G. Barto, and others, *Reinforcement learning: An introduction*, vol. 1. MIT press Cambridge, 1998.

[9] J. Gatheral, T. Jaisson, and M. Rosenbaum, "Volatility is rough," *Quantitative finance*, vol. 18, no. 6, pp. 933–949, 2018.

[10] P. Guasoni and M. Rásonyi, "Hedging, arbitrage and optimality with superlinear frictions," 2015.

Usage of AI tools

- ChatGPT – Research and review
- Cursor – Programming