Robust Risk-Aware Reinforcement Learning (RRA-RL) Applied in Dynamic Asset Allocation

USC Viterbi School of Engineering

Sahana Ramesh, Yiheng Lu, Shaswat Srivastava, Ziwei Duan, Jun Jie Ou Yang

Dept. of Computer Science, University of Southern California

Motivation

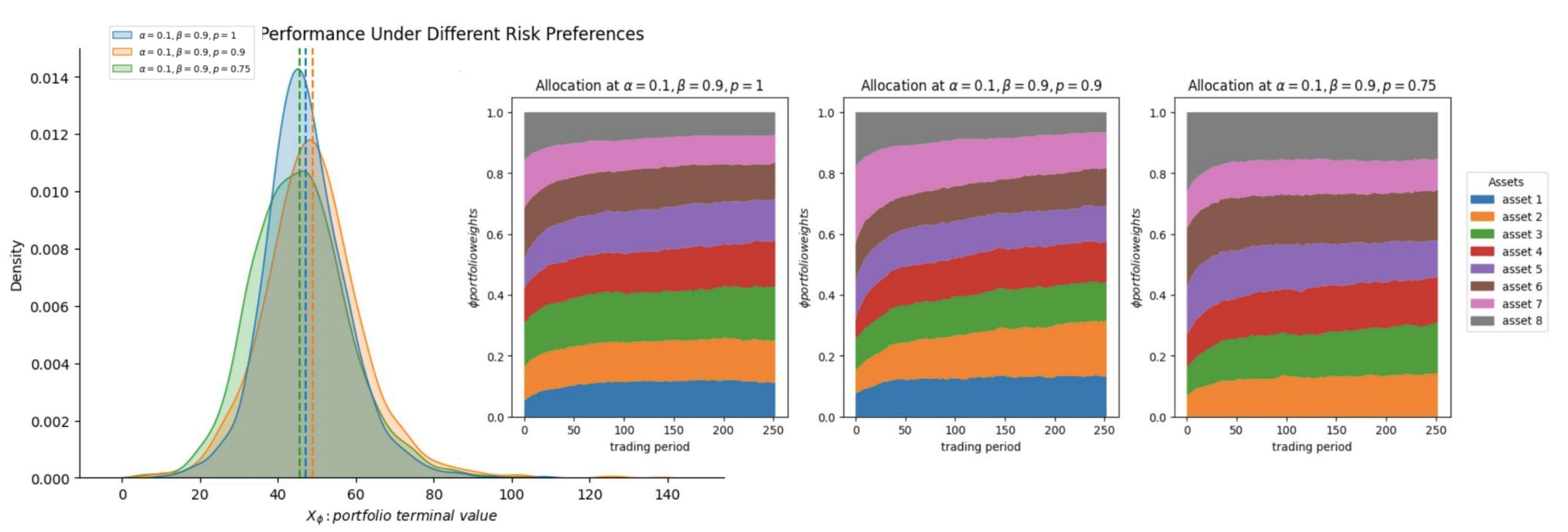
- **Dynamic asset allocation** can be cast in the form of a stochastic optimization problem.
- Finding the *optimal policy incorporating risk* (uncertainty) inherent in a model is difficult yet of paramount importance.
- RRA-RL framework presented in Jaimungal et al. [1] claims to be a *generic risk-aware RL framework* applicable to a variety of financial mathematics problems.

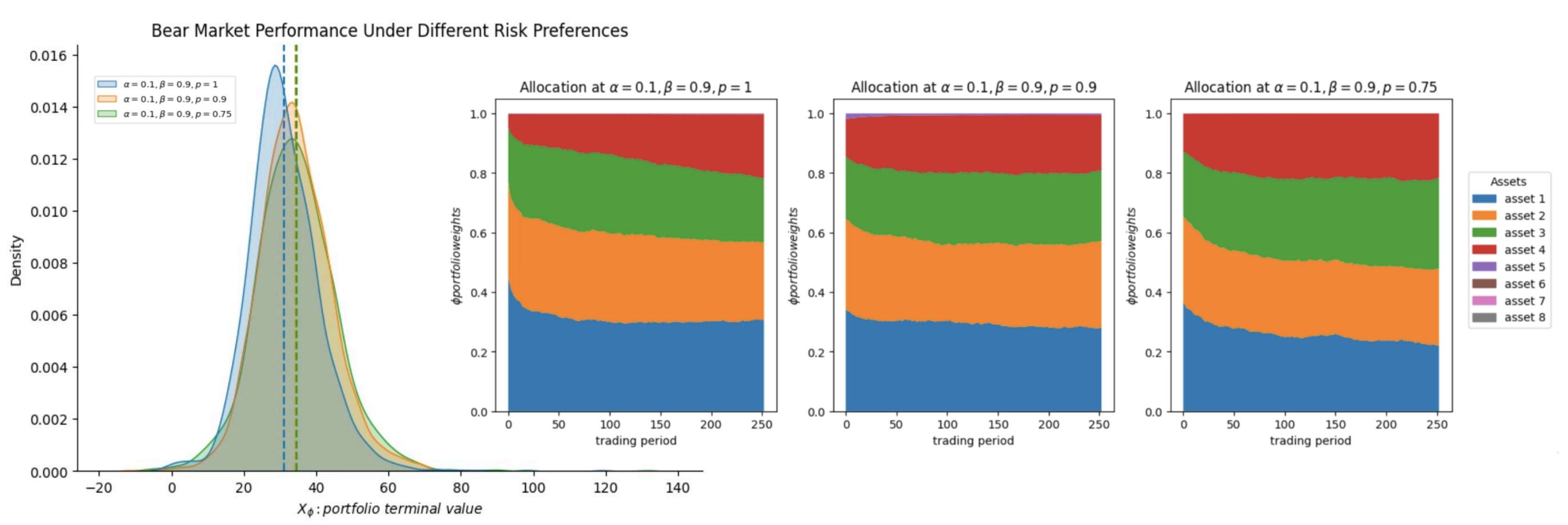
Objectives

- Extend application of a generic risk-aware RL framework to optimize **dynamic asset allocation**.
  - To model price uncertainty, generate simulated price paths that auto-regress, as well as sample from real market data.
- Compare financial performance of **RRA-RL** against **five RL techniques** that each has a unique way to mitigate risk:
  - Proximal Policy Optimization (PPO)
  - Advantage Actor Critic (A2C)
  - Deep Deterministic Policy Gradients (DDPG)
  - Soft Actor Critic (SAC)
  - Twin Delayed DDPG (TD3)

Results

- Trained on simulated asset prices with returns and variances that rank order.
- Compared density of terminal wealth and resulting asset allocation under **varying risk preferences p** during bull market (above) and bear market (below).
  - p is a hyper-parameter ∈ [0, 1] that allows an S-shape utility.
  - Higher p indicates the investor is more loss-avoiding.

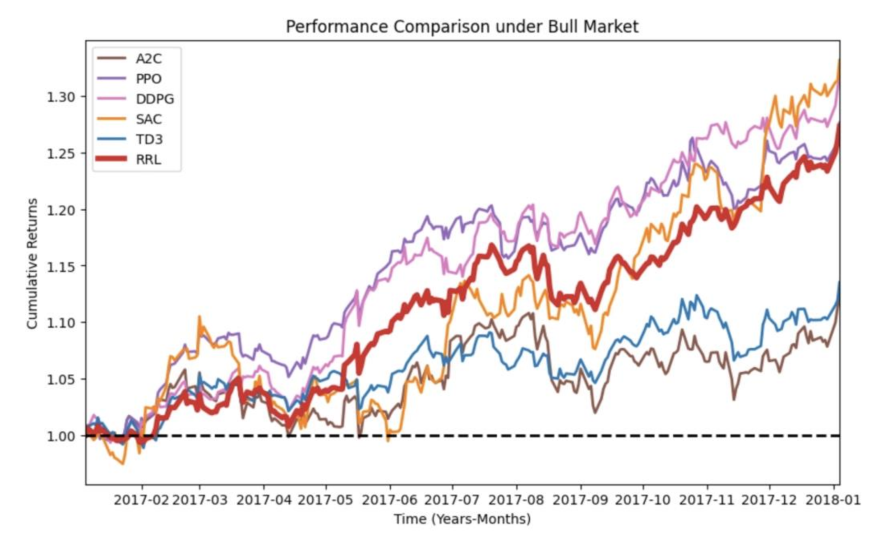


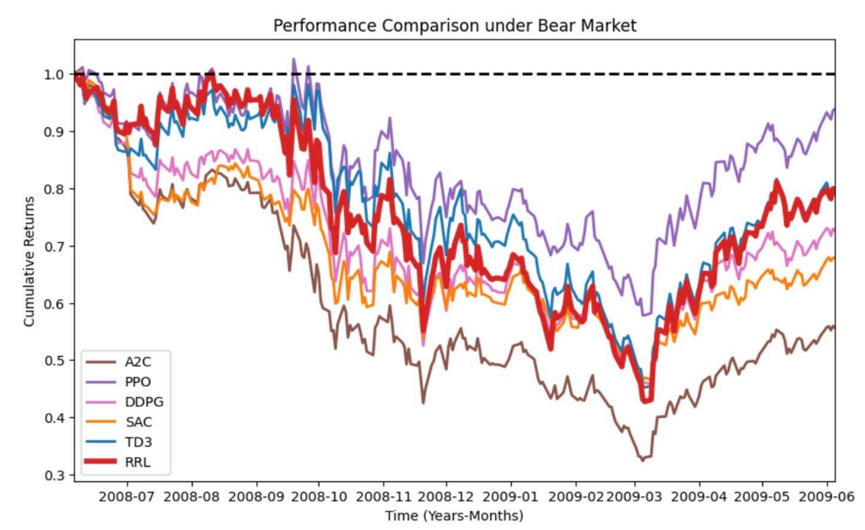


Benchmark & Comparison

Bull vs Bear Markets:

- Comparison focus on the *best* and *worst* of market scenarios.
- This makes the **tail ratio** especially useful for comparison.



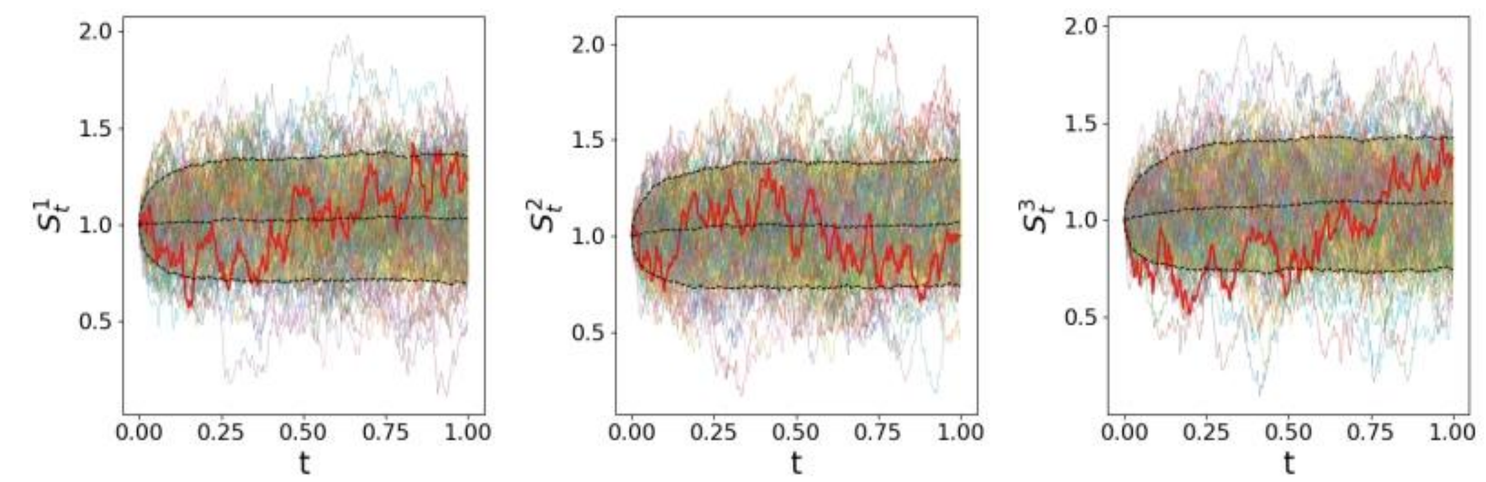


Performance of RRA-RL is consistently in the **top 3** in both the bull and bear market scenarios. More comparison can be conducted with *varying risk preferences*.

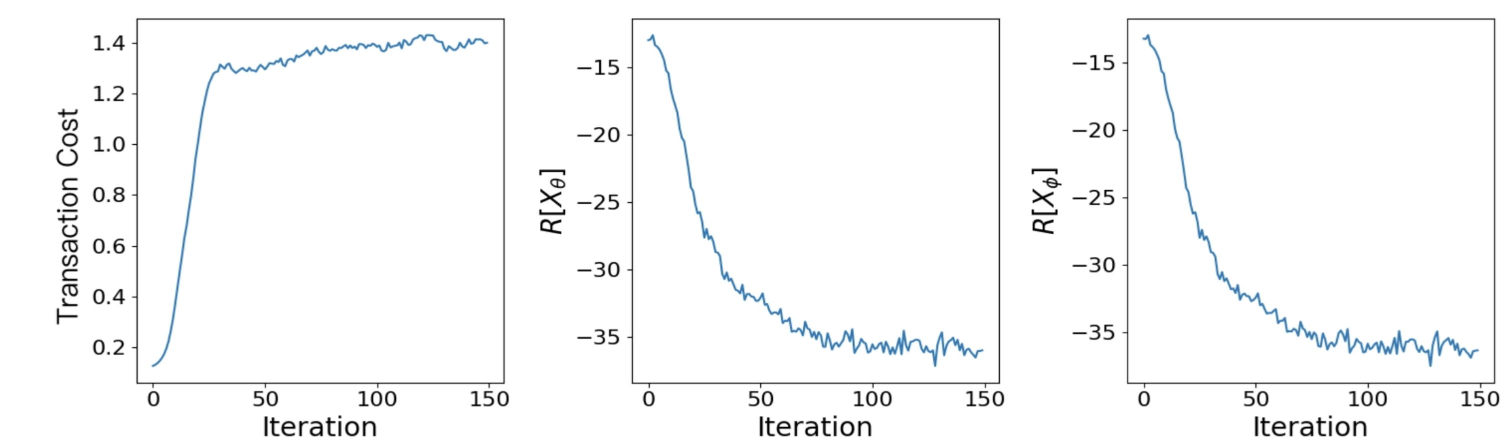
		Annual Returns	Annual Volatility	Sharpe Ratio	Calmar Ratio	Stability	Tail Ratio	Sortino ratio
Bull Market	A2C	12.4	12.7	0.99	1.56	0.50	0.98	1.43
	PPO	27.7	10.0	2.50	5.56	0.90	1.29	4.03
	DDPG	32.0	10.8	2.62	8.56	0.95	1.20	4.11
	TD3	13.5	9.6	1.37	2.63	0.80	1.09	2.01
	SAC	33.2	15.1	1.97	3.32	0.81	1.33	3.25
	RRL	27.3	9.5	2.58	5.59	0.94	1.20	4.04
Bear Market	A2C	-44.4	53.9	-0.82	-0.68	0.66	0.88	-1.14
	PPO	-6.1	53.5	0.15	-0.14	0.30	0.97	0.22
	DDPG	-27.7	55.3	-0.30	-0.51	0.50	0.95	-0.43
	TD3	-20.0	59.3	-0.08	-0.37	0.47	1.01	-0.12
	SAC	-32.1	47.5	-0.58	-0.60	0.59	0.87	-0.81
	RRL	-20.5	62.0	-0.06	-0.36	0.46	0.99	-0.09

Implementation

- **Input S** = $(S_0, \ldots S_i, \ldots, S_{T-1})$ denotes the price path of assets in the portfolio at trading time points $0 < t_1 < t_2 < \cdots < t_{T-1}$.
- **Output** $\varphi = [w_1, w_2, \ldots, w_n]^{(i)}$ denotes a policy, which is the weight of each asset in the portfolio at the sequence of trading time points.
- Our agent aims to find the optimal policy that **minimizes the maximum risk** of the terminal portfolio value, $X\varphi$.
  - Inner problem *(maximization)* results in a robust estimate of the risk of $X\varphi$ *(the worst outcome)*.
  - Outer problem *(minimization)* seeks to find the *best dynamic asset allocation of the worst outcome*, and therefore the optimal policy returned is risk-aware and robust to price uncertainty.



Simulated price paths (above), **Convergence** of value of optimal policy $X\varphi$ (below)



References

[1] S. Jaimungal, S. Pesenti, Y. S. Wang, and H. Tatsat, “Robust risk-aware reinforcement learning,” arXiv.org, https://arxiv.org/abs/2108.10403 (accessed Nov. 7, 2023).