Risk-aware multi-armed bandit problem with application to portfolio selection

Xiaoguang Huo and Feng Fu

PRESENTED BY

Jamal Verdiyev

What Are Multi-Armed Bandits (MAB)?

There are K slot machines, each with an unknown average reward μ_i

At each step, the player chooses one machine → receives a random reward.

Goal: maximize the total reward over N rounds.

The Exploration–Exploitation Dilemma

Regret measures how much reward you lost because you didn’t always play the best arm from the start.

The dilemma:

Exploitation: keep playing the best-known machine

Exploration: try others in case a better one exists

How to Balance Exploration and Exploitation?

ε-greedy:

regret grows linearly

with probability ε, choose randomly
otherwise(1-ε), pick the best

How to Balance Exploration and Exploitation?

UCB1

regret grows with O(logn)

the main idea: what if the machine we haven’t tried much is actually better, we just do not have enough information about it

selects the machine with the highest upper confidence bound:

$$\mathrm{UCB}_i = \bar{R}_i + \sqrt{\frac{2\log t}{T_i(t)}}$$

uncertainty measure

Why Is This Not Enough for Finance?

In portfolio investment, risk is as important as return.

"A more important variant is the risk-aware setting, where the learner considers risk in the objective instead of simply maximizing the cumulative reward."

Risk-Aware Methods

VaR (Value-at-Risk): maximum loss with probability β
CVaR (Conditional VaR): average loss in the worst (1−β)·100% of cases

CVaR

a coherent risk measure (subadditive → diversification helps)

$$\mathrm{CVaR}_\alpha(X+Y) \leq \mathrm{CVaR}_\alpha(X) + \mathrm{CVaR}_\alpha(Y)$$

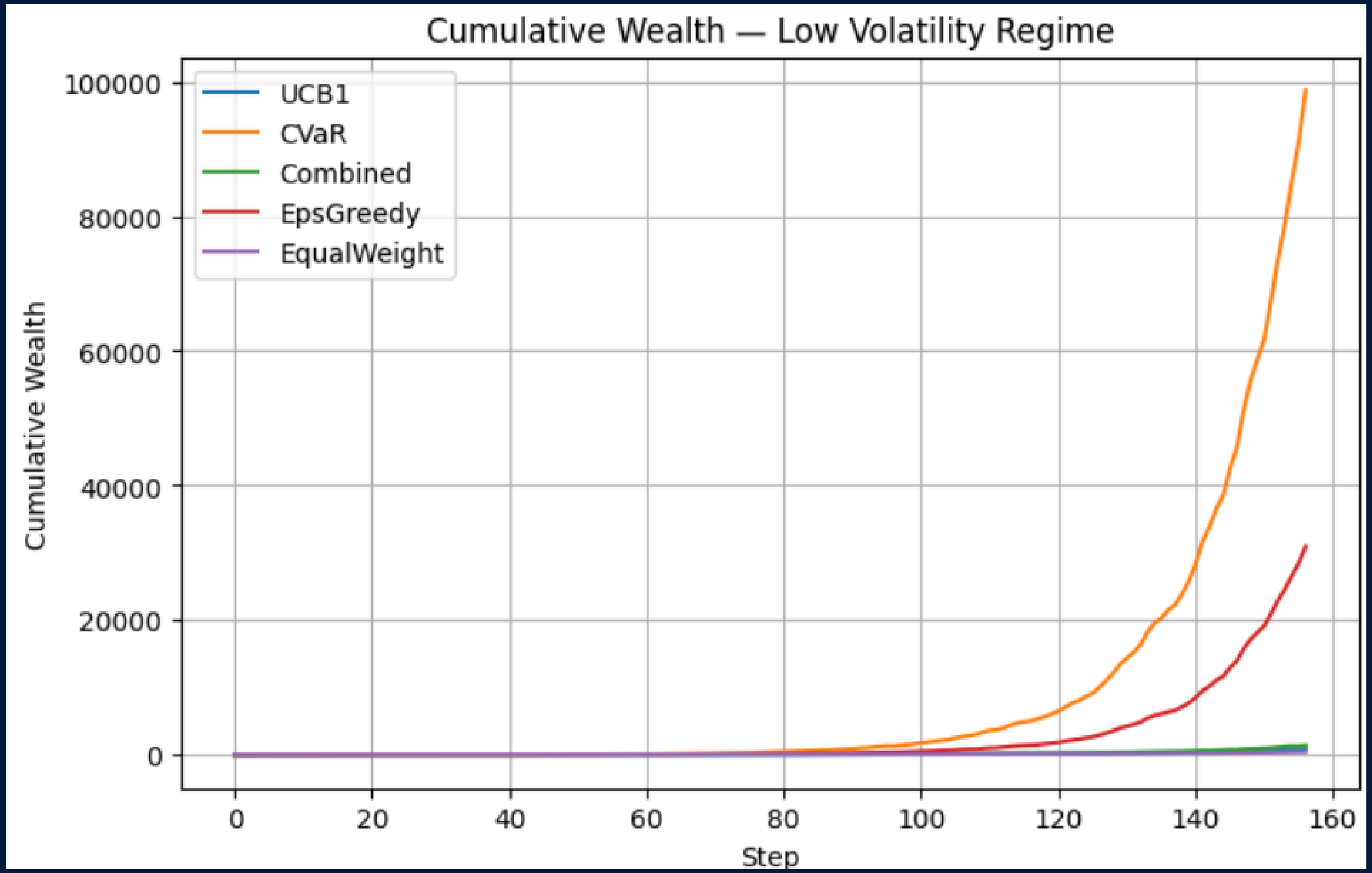
VaR

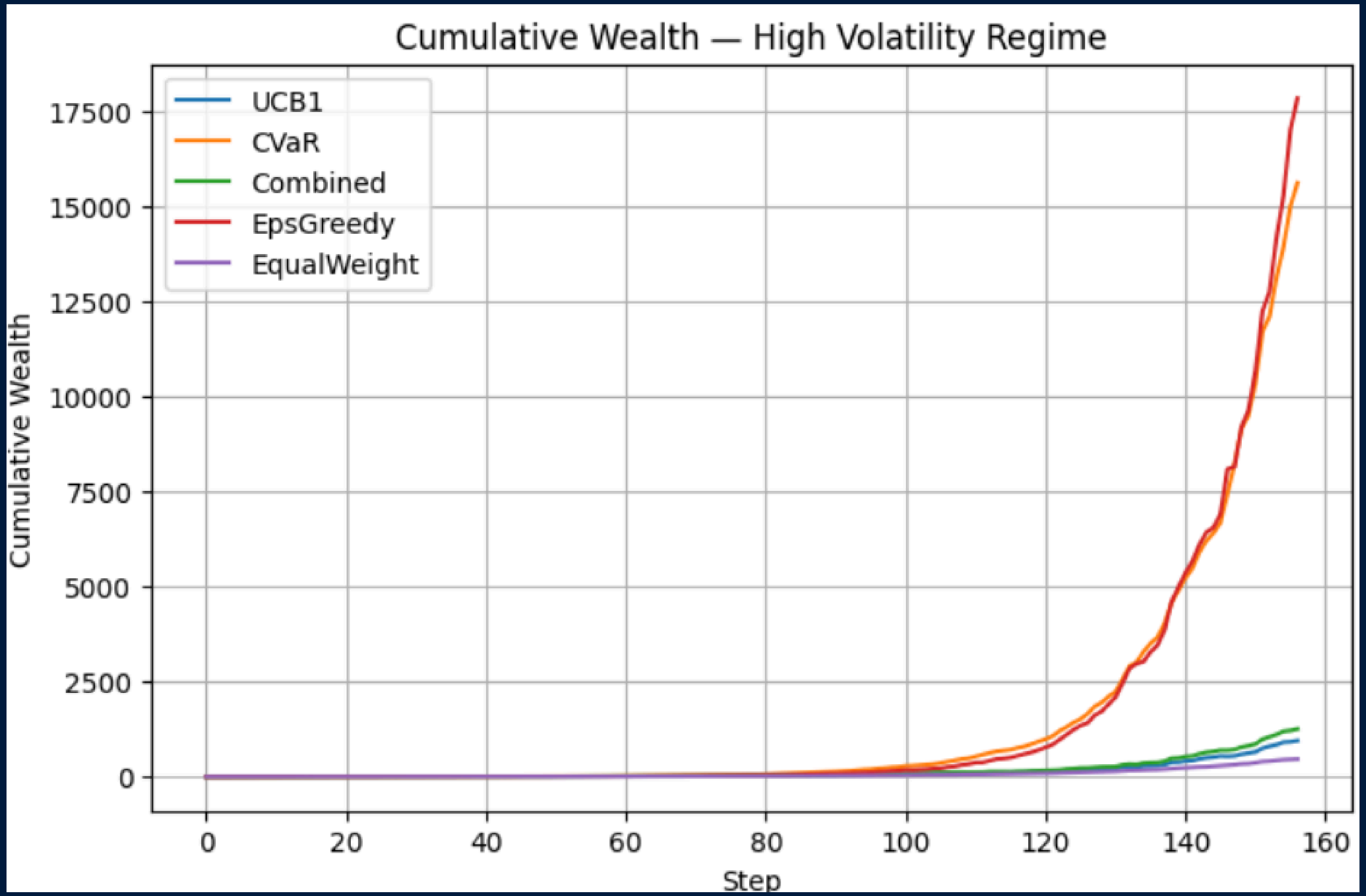
not subadditive → can encourage concentrated risk

Equal-Weighted Portfolio as a Baseline

Simplest diversification approach
assign equal weights to all selected assets

$$w_i = \frac{1}{K}, \quad i = 1, 2, \ldots, K$$


Cumulative Wealth — Low Volatility Regime
UCB1
CVaR
Combined
EpsGreedy
EqualWeight
Cumulative Wealth
100000
80000
60000
40000
20000
0
0
20
40
60
80
100
120
140
160
Step


Cumulative Wealth — High Volatility Regime
UCB1
CVaR
Combined
EpsGreedy
EqualWeight
Cumulative Wealth
17500
15000
12500
10000
7500
5000
2500
0
0
20
40
60
80
100
120
140
160
Step

Why Classical Approaches Fail?

Classical bandits maximize μ_i(return), ignoring risk
Risk-aware bandits select a single asset, not a portfolio

Portfolio investing requires:

Choosing a set of assets,
Assigning weights,
Accounting for correlations and systemic risk

Combined sequential portfolio selection algorithm

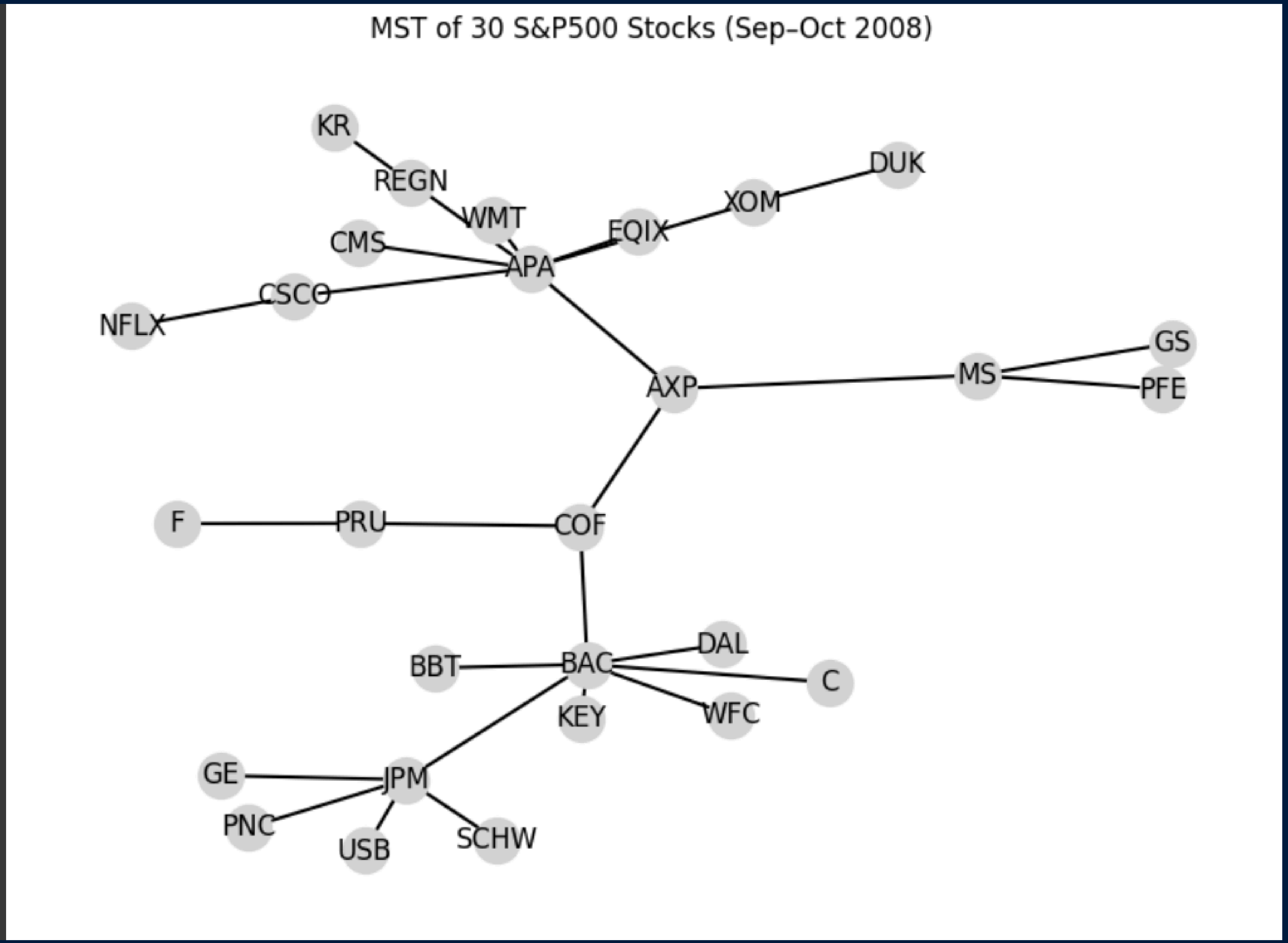
Build a correlation matrix from historical data

Convert it to a distance metric using

$$d_{i,j} = \sqrt{2(1-\rho_{i,j})}$$

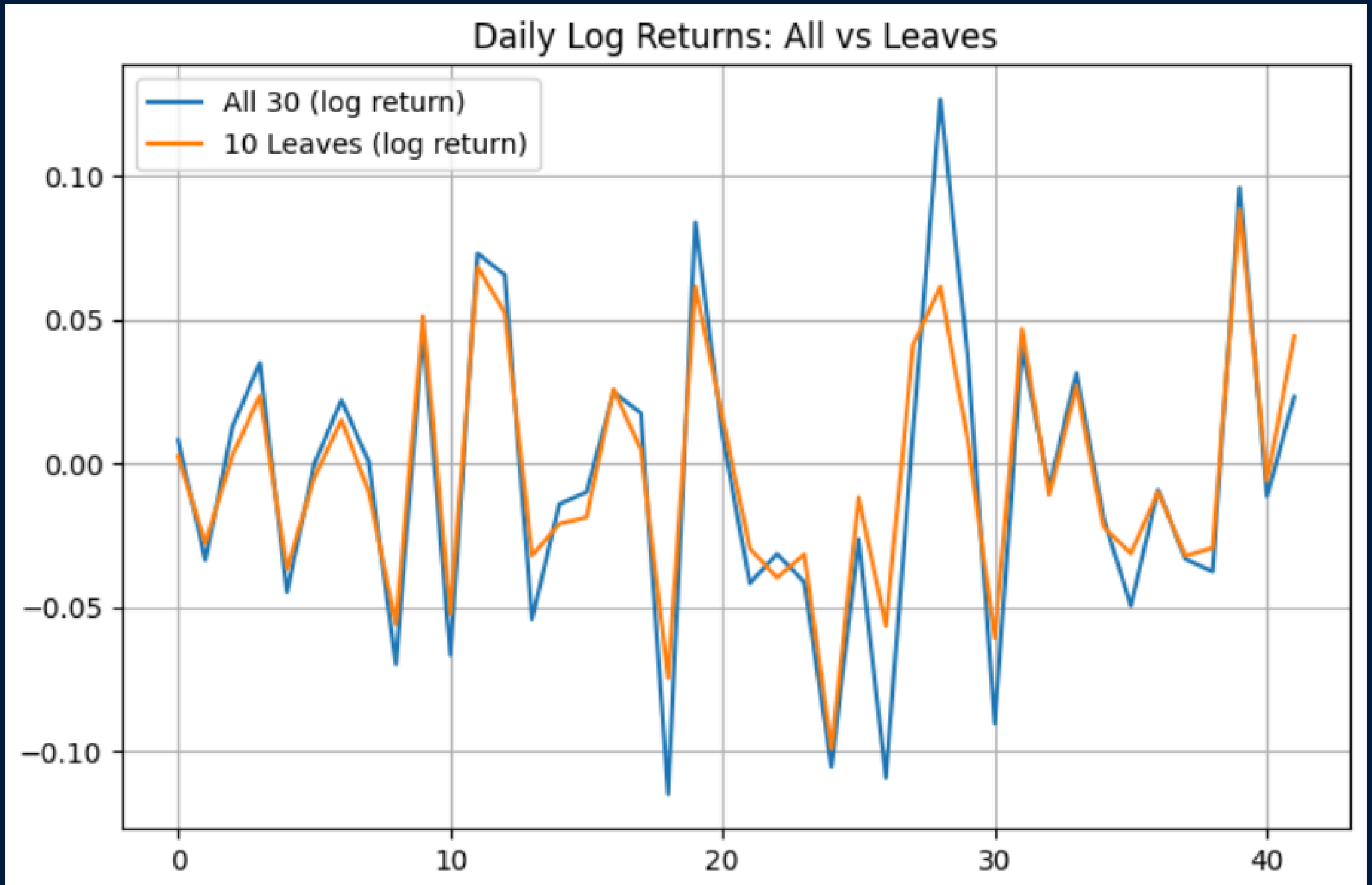
Construct a Minimum Spanning Tree

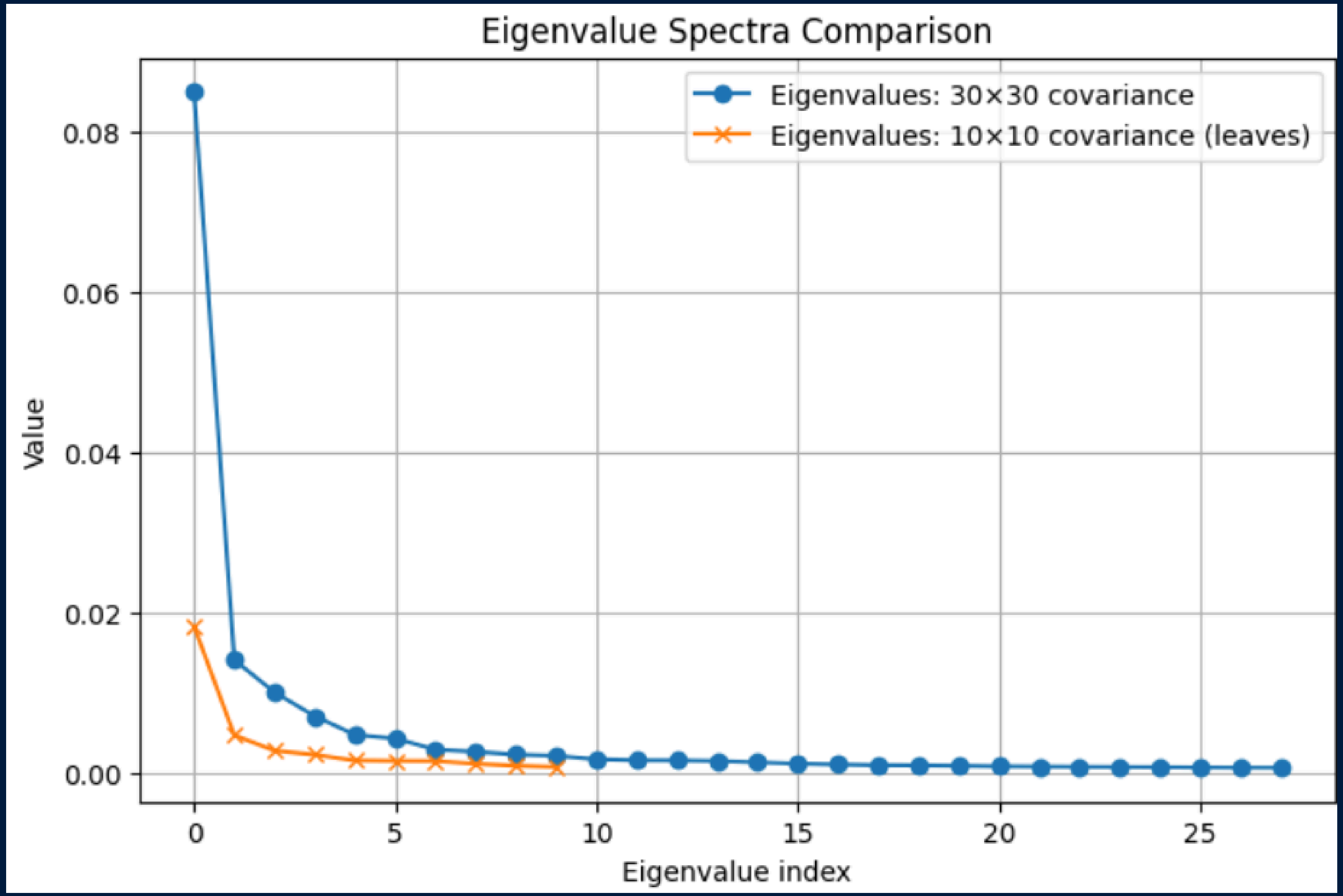
Select peripheral (leaf) assets, less exposed to systemic risk



MST of 30 S&P500 Stocks (Sep–Oct 2008)

Combined sequential portfolio selection algorithm





Combined sequential portfolio selection algorithm

Return-focused (UCB1):

$$\omega_t^M = e_{I_t^*}, \quad I_t^* = \arg\max_i \left(\bar{R}_i(t) + \sqrt{\frac{2\log t}{T_i(t)}} \right)$$

Risk-focused (CVaR):

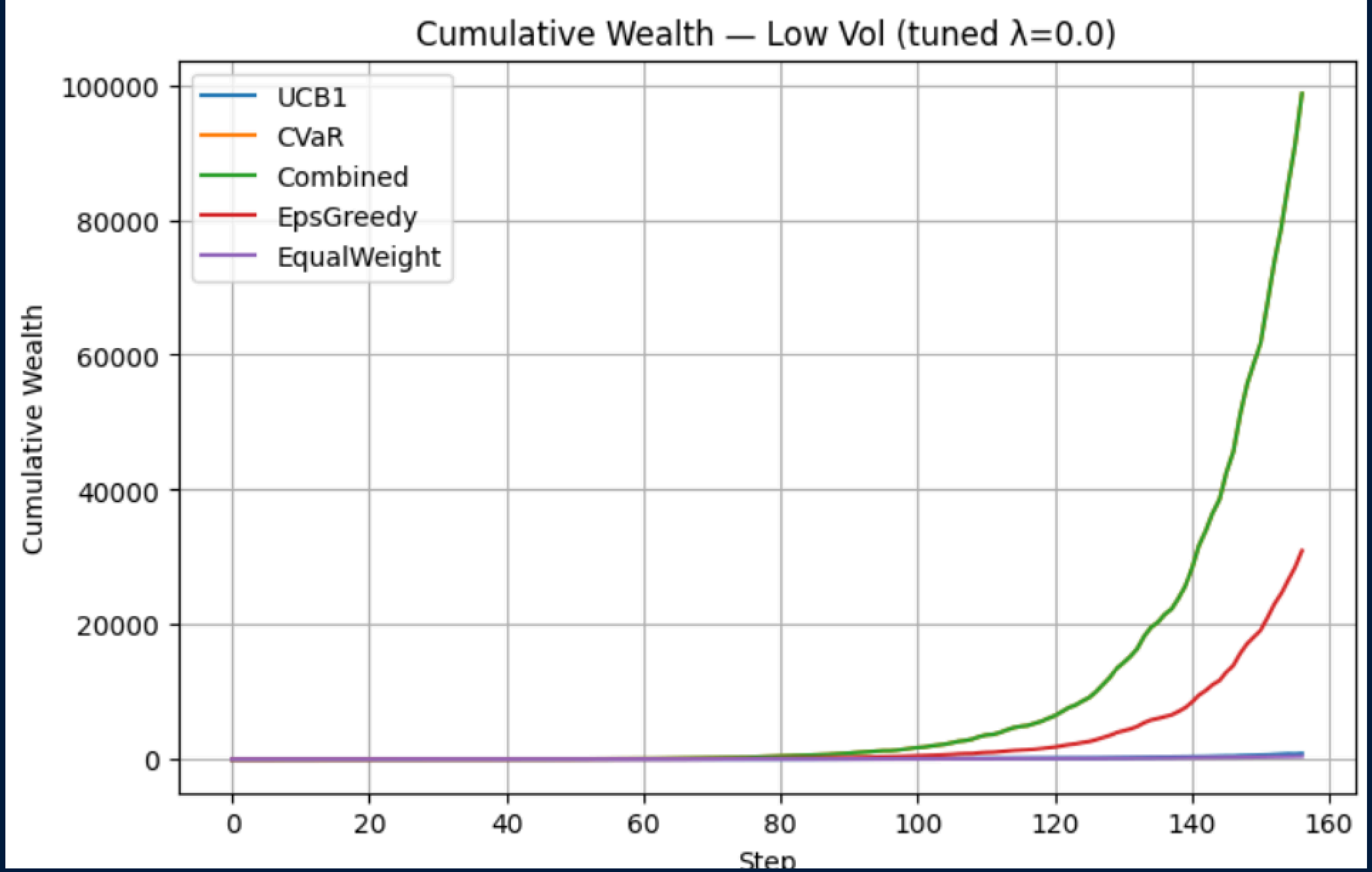
$$\omega_t^C = \arg \min_{(u,\alpha)\in W\times\mathbb{R}} \tilde{F}_\gamma(u, \alpha, t)$$

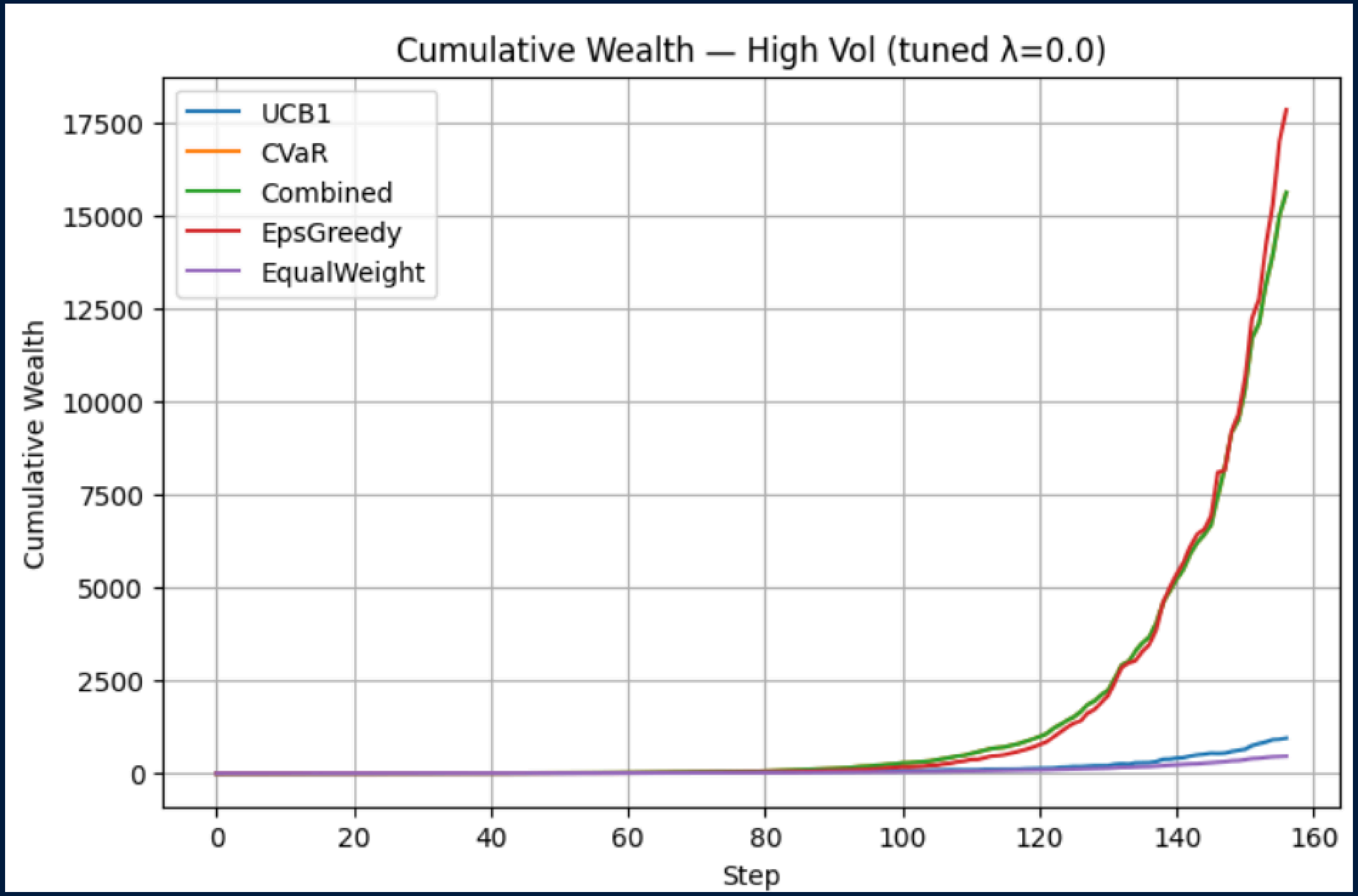
Final portfolio:

$$\omega_t^* = \lambda\omega_t^M + (1-\lambda)\omega_t^C$$

$\lambda \in [0,1]$ controls the degree of risk-taking

Simulation Results





Thank you for your attention!