Rethinking Risky Portfolio Allocation as An Online Mixed Integer Stochastic Bandit Problem

Jeffrey D. Varner, R.F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY 14850

Methods

Utility Function: Cobb-Douglas utility measures investor satisfaction:

$$U(n_1, \ldots, n_P) = \kappa(\gamma) \prod_{i \in \mathcal{P}} n_i^{\gamma_i}$$

where γ_i is the preference for asset i (output of a preference model), $\kappa(\gamma)$ scales the function, and n_i is the number of shares (what we compute).

Investor preference model: The γ_i preference coefficients can reflect market conditions, sentiment, and other asset-specific information through a feature vector $\mathbf{x}_i \in \mathbb{R}^m$:

$$\gamma_i = \sigma\left(\mathbf{x}_i^\top \theta_i\right) \quad \forall i \in \mathcal{P}$$

where $\sigma : \mathbb{R} \to \mathbb{R}$ is an activation function $\sigma_\theta(x) \in [-1, 1]$, and $\theta_i \in \mathbb{R}^p$ $(p = m + 1)$ denotes the feature weights (and bias), learned from data or set based on subjective beliefs.

Allocate investment budget B across portfolio assets, where p_i is the price of asset i. The optimal portfolio is the solution of the utility maximization problem:

$$\begin{aligned}
\underset{n_1,\ldots,n_P}{\text{maximize}} \quad & \kappa(\gamma) \prod_{i \in \mathcal{P}} n_i^{\gamma_i} \\
\text{subject to} \quad & B = \sum_{i \in \mathcal{P}} n_i \, p_i \\
& \epsilon \, a_i \leq n_i \leq a_i \left(\frac{B}{p_i}\right) \quad \forall i \in \mathcal{P} \\
& a_i \in \{0, 1\} \quad \forall i \in \mathcal{P} \\
& \epsilon \in \mathbb{R}_+ \quad \text{(hyperparameter)}
\end{aligned}$$

where $a_i = \{0, 1\}$ indicates asset availability (bandit arm)

The share optimization problem has an **analytical solution**. Let $S = \{i \mid a_i = 1\}$ be the available assets; $S_+ = \{i \mid \gamma_i > 0\}$ the preferred ones, $S_- = \{i \mid \gamma_i < 0\}$ the non-preferred. The optimal shares n_i^* are:

$$n_i^* = \begin{cases} \left(\dfrac{\gamma_i}{\sum_{j \in S_+} \gamma_j}\right) \dfrac{B - \epsilon \sum_{k \in S_-} p_k}{p_i} & \forall i \in S_+ \\ \epsilon & \forall i \in S_- \end{cases} \quad \blacksquare$$

Here, n_i^* is the optimal number of shares for asset i, balancing preferences and budget constraints.

Bandit

Results

Portfolio simulations for the bandit investor model. **A**: Example K = 10 asset portfolio with random fill prices and $\lambda = 1.0$. The bandit portfolio (red) is near the efficient frontier, but not on it. **B**: Efficient frontiers for different λ values, showing trade-offs between risk and return. Out-of-sample performance over 182 trading days (2025), comparing bandit, maximum Sharpe ratio, and SPY index portfolios (K=20) at varying risk aversion: no risk ($\lambda = 0$, **C**), moderate ($\lambda = 1.0$, **D**), high ($\lambda = 2.0$, **E**).

GitHub