

Structure-Aware Distributionally Robust Portfolio Optimization

진영봉

FRE LAB

Jul. 14, 2026

Background

- Optimization under Uncertainty

Phase 1: Stochastic Programming (SP) [1,2]

- 가정: 과거 데이터로 미래의 확률 분포를 완벽히 추정할 수 있음
- 방식: 관측된 경험적 분포 ($\hat{\mathbb{P}}_N$)나 가정한 모수적 분포를 참이라고 믿고 기대 수익을 극대화(혹은 위험 최소화)
- 한계: 과거 데이터가 미래를 대변하지 못하기 때문에 입력 변수 (기대수익률, 공분산)의 미세한 오차가 최적화 과정에 증폭 → out-of-sample 성능 하락

Phase 2: Robust Optimization (RO) [3-5]

- 가정: 분포를 정확하게 추정하는 것은 불가능 → 변수가 존재할 수 있는 범위 (uncertainty set)는 알 수 있음
- 방식: 불확실성 집합 $\mathcal{U}$ 를 정의한 뒤, 그 안에서 발생할 수 있는 최악의 시나리오를 방어

$$\min_{w \in \mathcal{W}} \sup_{\xi \in \mathcal{U}} L(w, \xi)$$

- 한계: 확률적 정보를 무시하기 때문에 현실에서 일어날 확률이 극히 희박한 상황까지 고려하게 됨 → 극단적으로 보수적

Phase 3: Distributionally Robust Optimization (DRO, Moment-based) [6]

- 가정: 정확한 분포는 모르지만, 평균이나 분산 같은 통계적 moments는 신뢰가능 → SP와 RP의 타협점
- 방식: 특정 평균 μ 와 공분산 Σ 를 만족하는 모든 가능한 확률 분포들의 집합 (모호성 집합, ambiguity set $\mathcal{P}$)를 만들고 그중 최악의 '분포'를 방어

$$\min_{w \in \mathcal{W}} \sup_{\mathbb{P} \in \mathcal{P}(\mu, \Sigma)} \mathbb{E}_{\mathbb{P}}[L(w, \xi)]$$

- 한계: 주어진 모멘트 제약만 만족하면 되기 때문에, 임의의 분포 형태가 허용됨 → worst-case 분포가 실제와 동떨어진 형태

Phase 4: data-driven / Wasserstein DRO [7-9]

- 가정: 데이터를 중심으로 하되, 주변의 확률적 오차를 거리 개념으로 통제
- 방식: 경험적 분포 $\hat{\mathbb{P}}_N$ 를 중심으로 반경 ϵ 내에 있는 분포들로 ambiguity set 구성
- Why Wasserstein?

KL-Divergence는 두 분포의 support가 일치해야 함

경험적 분포는 이산적인 포인트의 집합이지만 실제 미래 분포는 연속적일 수 있음

Literature Review

- Notation

- $x_i \in \mathbb{R}^n$: 관측된 수익률 데이터 / True 분포에서 나온 데이터
- $\hat{P}_N := \frac{1}{N} \sum_{i=1}^N \delta_{\hat{x}_i}$: 경험적 분포
- $\xi_i \in \mathbb{R}^n$: 가상 시나리오 $Q \in \mathcal{U}_\epsilon$ 에서의 확률 변수
- $\epsilon > 0$: 교란 허용 반경

- WDRO - Esfahani & Kuhn (2018) [8]

Step 1: ambiguity set 정의

- $\mathcal{U}_\epsilon(\hat{P}_N) = \{Q: W_c(\hat{P}_N, Q) \leq \epsilon\}$, $W_c(\hat{P}_N, Q) = \inf_{\pi \in \Pi(\hat{P}_N, Q)} \mathbb{E}_\pi[c(x, \xi)] = \inf_{\pi \in \Pi(\hat{P}_N, Q)} \int c(x, \xi) d\pi(x, \xi)$
- 관측 데이터 x_i 를 가상 시나리오 ξ 로 이동시키는 최소 비용이 ϵ 이하인 모든 분포 Q 의 집합
- 기본 설정: $c(x, \xi) = \|x - \xi\|$

Step 2: Solving worst-case expectation problems

- $\sup_{Q \in \mathcal{U}_\epsilon(\hat{P}_N)} \mathbb{E}_Q[\ell(\xi)], \ell(\xi) := \max_{k \leq K} \ell_k(\xi)$
- K 개의 기초함수들의 pointwise maximum으로 표현되어야 함
- 이 문제를 원래 분포 Q 에 대해 직접 접근하는 것은 무한차원 최적화이므로 해결하기 어려움

Literature Review

- WDRO - Esfahani & Kuhn (2018) [8]

Step 3: Tractable Reformulation

- $\sup_{Q \in \mathcal{U}_\epsilon(\hat{P}_N)} \mathbb{E}_Q[\ell(\xi)] = \sup_Q \inf_{\lambda \geq 0} \{ \mathbb{E}_Q[\ell(\xi)] + \lambda(\epsilon - W_c(\hat{P}_N, Q)) \} = \inf_{\lambda \geq 0} \{ \lambda\epsilon + \sup_Q [\mathbb{E}_Q[\ell(\xi)] - \lambda W_c(\hat{P}_N, Q)] \} = \inf_{\lambda \geq 0} \left\{ \lambda\epsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi} [\ell(\xi) - \lambda c(x_i, \xi)] \right\}$
- Lagrangian $\rightarrow$ Strong Duality $\rightarrow$ Inner $\sup_Q$
- 무한 차원 분포 최적화 $\rightarrow$ 관측치 데이터를 기준으로 하는 유한차원 worst-case 문제의 합

Step 4: Example – Mean-risk portfolio optimization (CVaR)

- $\sup_Q \mathbb{E}_Q[\ell(\xi)] = \sup_Q \inf_{w \in W} \{ \mathbb{E}_Q[-w^\top \xi] + \rho \cdot \text{CVaR}_{\alpha, Q}(-w^\top \xi) \}$

$$= \sup_Q \inf_{w \in W} \left\{ \mathbb{E}_Q[-w^\top \xi] + \rho \inf_{\tau \in \mathbb{R}} \mathbb{E}_Q \left[\tau + \frac{1}{\alpha} \max\{-w^\top \xi - \tau, 0\} \right] \right\} \text{ by [10]}$$

$$= \sup_Q \inf_{w \in W, \tau \in \mathbb{R}} \mathbb{E}_Q \left[\max\{-w^\top \xi + \rho\tau, -\left(1 + \frac{\rho}{\alpha}\right)w^\top \xi + \rho\left(1 - \frac{1}{\alpha}\right)\tau \} \right]$$

$$= \inf_{w \in W, \tau \in \mathbb{R}} \sup_Q \mathbb{E}_Q \left[\max_{k \leq K} a_k \cdot w^\top \xi + b_k \cdot \tau \right]$$
 - Loss가 $K = 2$ 인 piecewise linear + convex 형태
 - ϵ 이 증가할수록 안전자산 비중이 늘어남
 - 포트폴리오 시뮬레이션에서 우수한 oos 성능 달성

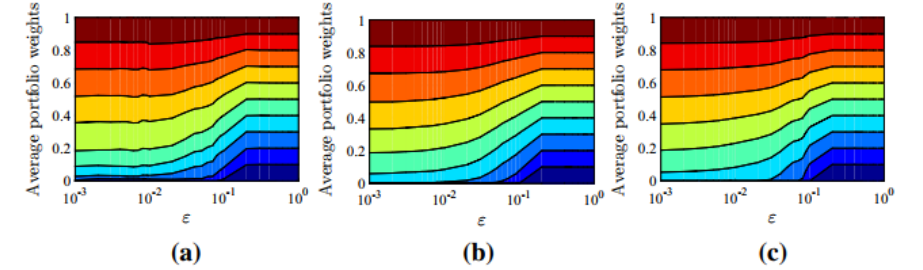


Fig. 4 Optimal portfolio composition as a function of the Wasserstein radius ϵ averaged over 200 simulations; the portfolio weights are depicted in ascending order, i.e., the weight of asset 1 at the bottom (dark blue area) and that of asset 10 at the top (dark red area). (a) $N = 30$ training samples. (b) $N = 300$ training samples. (c) $N = 3000$ training samples (color figure online)

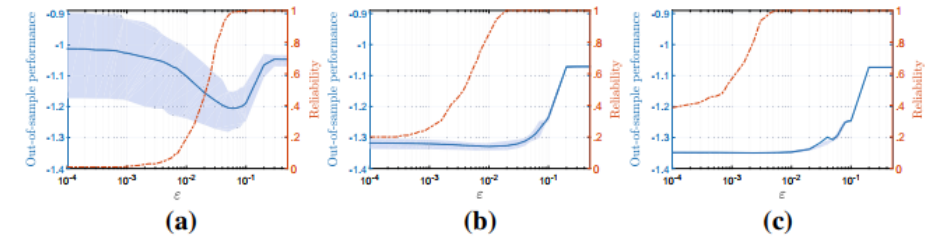


Fig. 5 Out-of-sample performance $J(\hat{x}_N(\epsilon))$ (left axis, solid line and shaded area) and reliability $\mathbb{P}^N[J(\hat{x}_N(\epsilon)) \leq \hat{J}_N(\epsilon)]$ (right axis, dashed line) as a function of the Wasserstein radius ϵ and estimated on the basis of 200 simulations. (a) $N = 30$ training samples. (b) $N = 300$ training samples. (c) $N = 3000$ training samples

Literature Review

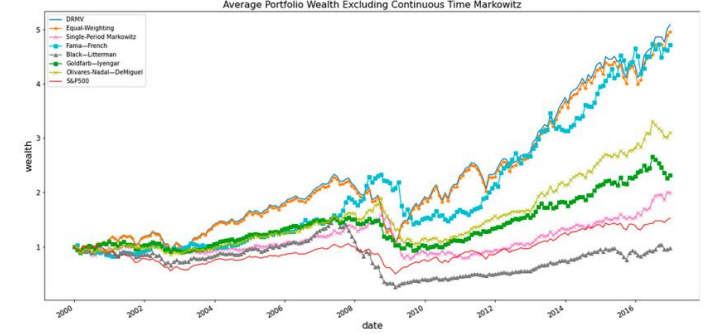
- WDRO – Blanchet et.al (2022) [9]

Step 1: Classical Markowitz portfolio selection problem

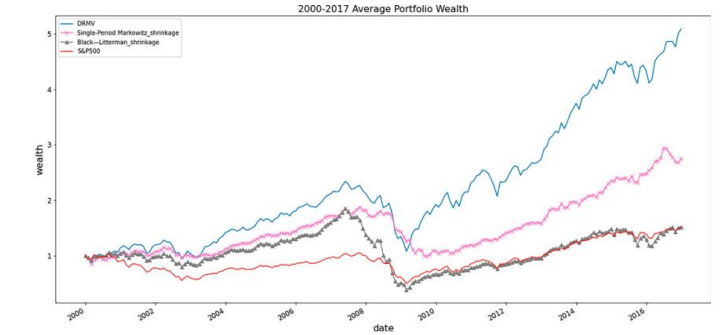
- $\min_w w^T \text{Var}(x) w \quad \text{s.t. } w^T \mathbf{1} = 1, w^T \mathbb{E}[x] = \rho$
- 목표 수익률을 ρ 정확하게 달성하는 모델 중 위험(분산)이 제일 작은 포트폴리오 선택

Step 2: Distributionally robust mean–variance (DRMV)

- 강건화 모델 (모호성 집합 Q 등장 $\rightarrow \xi$)
- $\min_{w \in \mathcal{F}_{\delta, \bar{a}}(N)} \max_{Q \in \mathcal{U}_{\delta}(\hat{P}_N)} w^T \text{Var}_Q(\xi) w, \quad \mathcal{F}_{\delta, \bar{a}} = \{w: w^T \mathbf{1} = 1, \min_{Q \in \mathcal{U}_{\delta}(\hat{P}_N)} \mathbb{E}_Q[w^T \xi] \geq \bar{a}\}$
- 목표 수익률의 등식 $\rightarrow$ 부등식 완화: 최악의 Q 에서도 최소 $\bar{a}$ 수익은 보장 ($\bar{a}$ 는 추후 ρ 통계적으로 역산하여 계산)
- 제약식 선형처리 (Append. A): $\min_{Q \in \mathcal{U}_{\delta}(\hat{P}_N)} \mathbb{E}_Q[w^T \xi] = w^T \bar{x} - \sqrt{\delta} \|w\|_p$
- 기대수익 α 고정 (Append. A): $\max_{Q \in \mathcal{U}_{\delta}(\hat{P}_N)} w^T \text{Var}_Q(\xi) w = \max_{\alpha \geq \bar{a}} \{ \max_{Q: \mathbb{E}_Q[w^T \xi] = \alpha} w^T \mathbb{E}_Q[\xi \xi^T] w - \alpha^2 \}, \mathbb{E}_Q[w^T \xi] = \alpha \geq \bar{a}$
- 목적식 duality (Append. B): $\max_{Q: \mathbb{E}_Q[w^T \xi] = \alpha} w^T \mathbb{E}_Q[\xi \xi^T] w = \inf_{\lambda_1 \geq 0, \lambda_2} \frac{1}{N} \sum_{i=1}^N \Phi(x_i) + \lambda_1 \delta + \lambda_2 \alpha, \quad \Phi(x_i) := \sup_{\xi} \{(w^T \xi)^2 - \lambda_1 c(x_i, \xi) - \lambda_2 w^T \xi\}$
- 최종 목적식 (Append. B): $\min_{w \in \mathcal{F}_{\delta, \bar{a}}(N)} \sqrt{w^T \text{Var}(x) w} + \sqrt{\delta} \|w\|_p \quad \text{s.t. } w^T \mathbf{1} = 1, w^T \bar{x} \geq \bar{a} + \sqrt{\delta} \|w\|_p$



Notes. All the portfolios except S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x-axis indicates the time in months (from 1 to 204), and the y-axis indicates the portfolio wealth. Initial wealth is set to be one.



Notes. All the portfolios except S&P 500 consist of 100 stocks, and the averages are calculated over 100 numerical experiments. The x-axis indicates the time in months (from 1 to 204), and the y-axis indicates the portfolio wealth. Initial wealth is set to be one.

Methodology

- 공통 문제:

분포를 알 수 없는 유한표본 의사결정

경험분포의 오차와 distribution shift 방어

불확실성의 크기와 형태 결정

- Esfahani & Kuhn (2018)

- 경험분포 주변 Wasserstein ambiguity set
- 유한표본 보장과 tractable reformulation
- ground cost (Euclidean), 사전 고정

- Blanchet et.al (2022)

- Wasserstein DR mean-variance portfolio
- 최악 평균·분산을 SOCP로 해결
- ground cost (Euclidean), 사전 고정

- **Research goal:** 시장의 내재적 정보를 학습하여 불확실성 집합 구조를 반영하는 SA-DRPO (Structure-Aware Distributionally Robust Portfolio Optimization)를 제안

EBM $p_\theta(x) \propto \exp(-E_\theta(x))$ 의 경험적 평균 에너지 곡률 M 을 WDRO의 metric으로 사용할 때,

목표 1: 이론적으로 정당화 되는지?

목표 2: 실증적으로 우수한 ambiguity set을 제공하는지?

Methodology

- Notation

- N : rolling-window 내 거래일 수 (5년 거래일; 252*5)
- d : 자산 수 (데이터 셋에 따라 각각 49; 100)
- B : 각 관측치에서 생성하는 Gaussian noise 수 (10)
- $\xi_i \in \mathbb{R}^n$: 가상 시나리오(자산 수익률)에서의 확률 변수

$$p_{\theta_t}(x) \propto \exp[-E_{\theta_t}(x)]. \quad s_{\theta_t}(x) = \nabla_x \log p_{\theta_t}(x) = -\nabla_x E_{\theta_t}(x) \in \mathbb{R}^d$$

- Step1. return distribution과 EBM 학습

- 직전 N 일의 d 개 자산 수익률 X_t 사용
- 수익률 자체를 예측하는 것이 아니라 분포의 국소적 구조를 나타내는 score를 학습

$$\mathcal{L}_{\text{DSM}}(\theta_t) = \frac{1}{NB} \sum_{i=1}^N \sum_{b=1}^B \left\| s_{\theta_t} \left(z_{t,i}^{(b)} \right) + \frac{\epsilon_{t,i}^{(b)}}{\sigma} \right\|_2^2 \quad \begin{aligned} z_{t,i}^{(b)} &= x_{t,i} + \sigma \epsilon_{t,i}^{(b)}, \\ \epsilon_{t,i}^{(b)} &\sim \mathcal{N}(0, I_d), \quad b = 1, \dots, B. \end{aligned}$$

- Step2. EBM score를 구조적 ground metric으로 변환

- 학습한 score의 squared magnitude를 집계하여 자산별 score-information을 계산
- Score-information이 큰 방향은 분포가 민감하게 변하는 방향으로 해석
- 이를 Euclidean metric 와 결합하여 안정적인 identity-anchored metric을 구성

$$\begin{aligned} \hat{h}_{j,t} &= \frac{1}{NB} \sum_{i=1}^N \sum_{b=1}^B \left[s_{\theta_t,j} \left(z_{t,i}^{(b)} \right) \right]^2 \\ D_t^{\text{EBM}} &= \text{diag} \left(\frac{d\hat{h}_{1,t}}{\sum_{\ell=1}^d \hat{h}_{\ell,t}}, \dots, \frac{d\hat{h}_{d,t}}{\sum_{\ell=1}^d \hat{h}_{\ell,t}} \right) \end{aligned}$$

$$M_t = (1 - \beta_t)I_d + \beta_t D_t^{\text{EBM}}, \quad 0 \leq \beta_t < 1.$$

Methodology

- Step3. 학습된 metric으로 ambiguity set 구성
 - step 2에서 얻은 M_t 를 transportation cost에 적용
 - M_t 값이 큰 방향은 분포 이동 비용이 커지고, 작은 방향은 상대적으로 이동이 쉬워짐
 - 따라서 동일한 radius에서도 EBM이 학습한 시장구조에 따라 ambiguity set의 모양이 달라짐

$$c_{M_t}(x, y) = (x - y)^\top M_t (x - y)$$

$$\mathcal{U}_t(\delta, M_t) = \left\{ P : \inf_{\pi \in \Pi(P, \hat{P}_{N,t})} \mathbb{E}_\pi[c_{M_t}(X, Y)] \leq \delta \right\}$$

- Step4. Distributionally robust primal problem
 - Step 3의 ambiguity set에서 발생할 수 있는 최악의 분포를 고려하여 portfolio weight를 결정
 - Wasserstein DRO primal problem은 다음과 같음

$$\min_{w \in \mathcal{F}_{\delta, \bar{\rho}_t}} \max_{Q \in \mathcal{U}_\delta(P_{N,t}; M_t)} w^\top \text{Var}_Q(\xi) w,$$

$$\mathcal{F}_{\delta, \bar{\rho}_t} = \left\{ w : \mathbf{1}^\top w = 1, w \geq 0, \min_{Q \in \mathcal{U}_\delta(P_{N,t}; M_t)} \mathbb{E}_Q[w^\top \xi] \geq \bar{\rho}_t \right\}.$$

Methodology

- Primal vs. Dual (Append C.)

$$\min_{w \in \mathcal{F}_{\delta, \bar{\rho}_t}} \max_{Q \in \mathcal{U}_\delta(P_{N,t}; M_t)} w^\top \text{Var}_Q(\xi) w,$$

$$\mathcal{F}_{\delta, \bar{\rho}_t} = \left\{ w : \mathbf{1}^\top w = 1, w \geq 0, \min_{Q \in \mathcal{U}_\delta(P_{N,t}; M_t)} \mathbb{E}_Q[w^\top \xi] \geq \bar{\rho}_t \right\}.$$



$$\min_w \sqrt{w^\top \widehat{\Sigma}_t w} + \eta q_{M_t}(w)$$

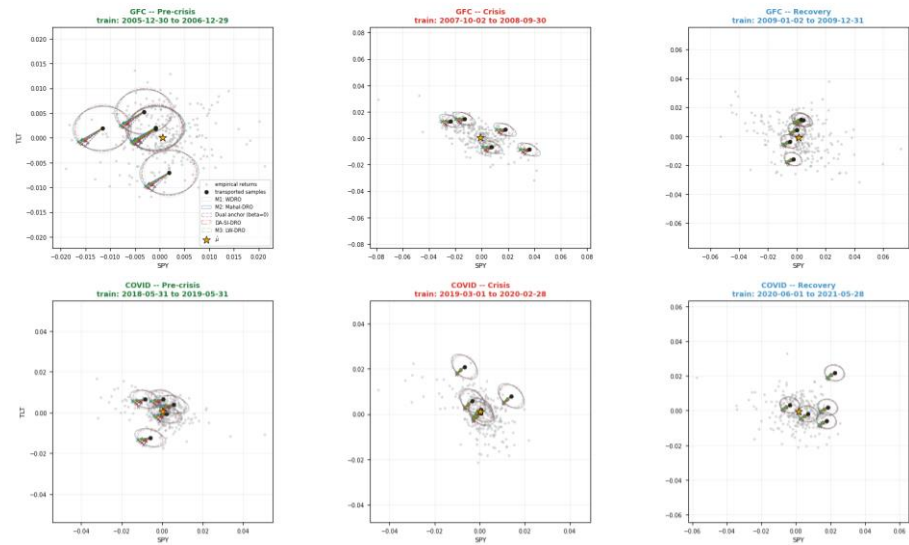
$$\text{s.t. } w^\top \bar{x}_t - \eta q_{M_t}(w) \geq \bar{\rho}_t,$$

$$\mathbf{1}^\top w = 1, \quad w \geq 0.$$

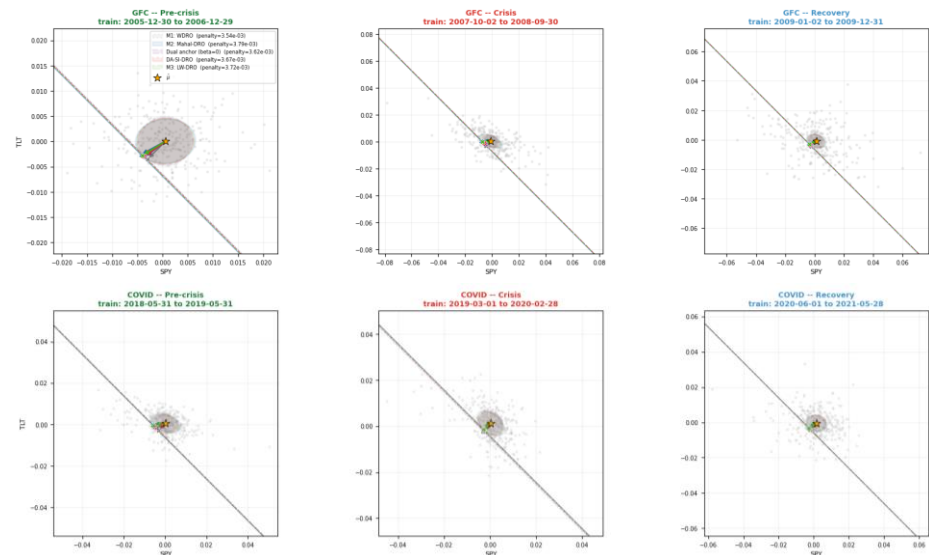
$$q_{M_t}(w) = \|w\|_{M_t^{-1}} = \sqrt{w^\top M_t^{-1} w}$$

$$M_t = (1 - \tilde{\beta}_t)I + \tilde{\beta}_t D_t^{\text{EBM}}$$

Primal witness: $T_h(x_i) = x_i + \tilde{\delta}_h^* x_i, |\tilde{\delta}_h^*|_h = \varepsilon$
fixed $w=[0.5, 0.5]$, $\text{epsilon}=5.0\text{e-}03$, $\text{lambda}=0.5$, $\text{beta}=0.5$



Dual support: $w^\top \tilde{\mu} - \varepsilon \sqrt{w^\top M^{-1} w}$
fixed $w=[0.5, 0.5]$, $\text{epsilon}=5.0\text{e-}03$, $\text{lambda}=0.5$, $\text{beta}=0.5$



Experimental results

- Data

Fama-French 49 industry portfolios

- Kenneth French Data Library의 49개 산업 포트폴리오 일별 수익률
- 산업 수준의 안정적인 횡단면 구조를 이용해 SA-DRPO의 structure-learning 효과 평가
- 개별 종목 잡음과 생존편향의 영향을 줄인 중간 차원 벤치마크
- 2010-01-05 ~ 2026-04-30

Random 100 Stocks from the S&P 500

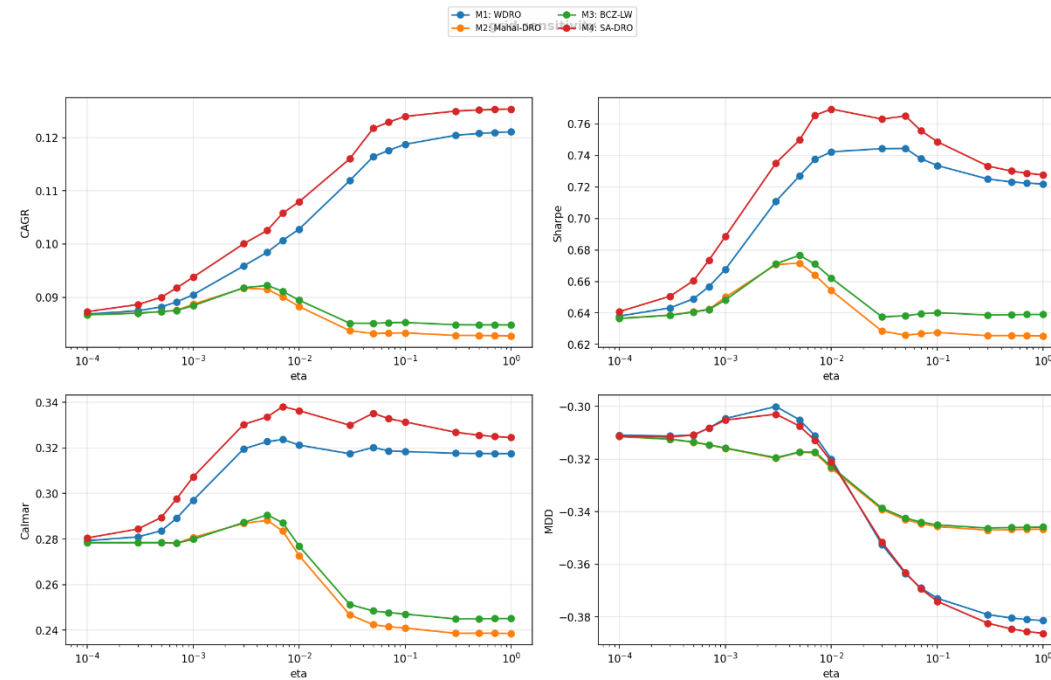
- S&P 500 구성 종목 중 100개 종목을 무작위 추출
- 자산 수가 많은 환경에서 공분산 및 분포 추정오차에 대한 강건성 평가
- 2010-01-05 ~ 2025-12-30

- Experimental setting

- 5년 train, 1달 test인 rolling window 방식 (single period rolling window)
- 비교 모형: M1: WDRO (Blanchet et.al (2022)), M2: 공분산 기반 DRO, M3: Ledoit-Wolf 기반 DRO, Equal-weight
- 평가 지표: Total return, CAGR, Sharpe, Calmar, MDD
- 수수료: 10bp

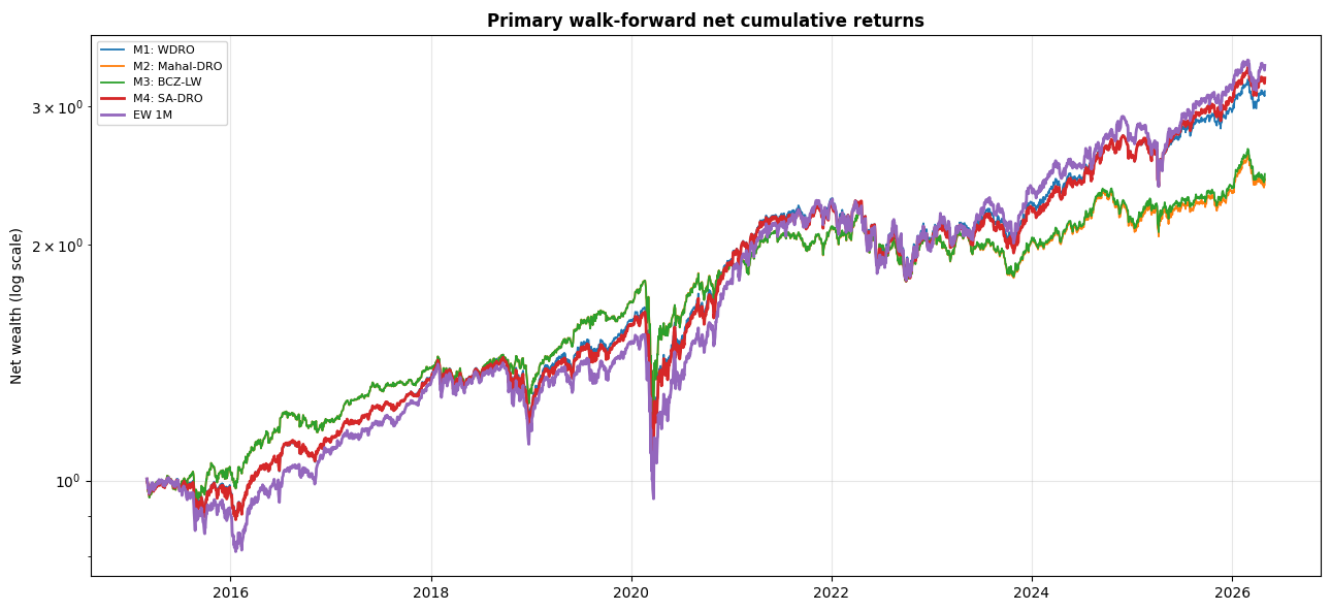
Experimental results

Fama-French 49 industry portfolios



Experimental results

Fama-French 49 industry portfolios



===== PRIMARY WALK-FORWARD OOS | start=2015-03-02 =====

	Total Return	CAGR	Sharpe	Sortino	Calmar	MDD	Turnover
Model							
M1: WDRO	+213.42%	+10.79%	0.721	0.871	0.306	-35.26%	9.58%
M2: Mahal-DRO	+142.29%	+8.26%	0.621	0.743	0.244	-33.92%	25.43%
M3: BCZ-LW	+145.81%	+8.40%	0.630	0.753	0.248	-33.87%	24.64%
M4: SA-DRO	+225.77%	+11.18%	0.739	0.897	0.318	-35.18%	10.67%
EW 1M	+238.07%	+11.55%	0.690	0.862	0.302	-38.25%	25.26%

conclusion

- Summary

- 기존 WDRO 는 isotropic ambiguity set을 사용하여 모든 방향의 분포 이동을 동일하게 취급
- 그러나 실제 금융 시장은 비등방적(asymmetric/anisotropic) 구조를 가지면, 위기 시 특정 방향으로 위험이 집중됨
- 본 연구는 EBM으로 return-space의 score를 학습하고, squared-score expectation으로부터 자산별 score-information을 추정
- 이를 이용해 시점별 anisotropic ground metric과 structure-aware Wasserstein ambiguity set을 구성
- SA-DRO는 기존 WDRO와 다른 데이터 기반의 robust penalty를 형성
- 예비 실험에서는 일부 파라미터 구간에서 기존 WDRO 대비 Sharpe ratio와 MDD 측면의 개선 가능성을 확인

- Future works

- 데이터 생존편향 문제
- 위기·고변동성 구간에서 학습된 score-information과 ambiguity set이 실제 위험 방향을 적절히 포착하는지 검증
- Epsilon 선택 시 단순 grid search가 아닌 coverage, realized tail loss 등의 기준으로 정교하게 calibration (or validation set 활용)
- 여러 S&P 500 random universes와 다양한 시장·기간에서 결과의 재현성을 검증

불확실성의 크기만 조절하는 기존 WDRO를 넘어, 시장 데이터로부터 불확실성의 방향 구조를 학습하고 이를 강건한 포트폴리오 의사결정에 반영하고자 함

References

1. Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77-91.
2. Dantzig, G. B. (1955). Linear Programming under Uncertainty. *Management Science*, 1(3/4), 197-206.
3. A. Ben-Tal, A. Nemirovski, (1998) Robust Convex Optimization. *Mathematics of Operations Research* 23(4):769-805.
4. Dimitris Bertsimas, Melvyn Sim, (2004) The Price of Robustness. *Operations Research* 52(1):35-53.
5. Fabozzi, F. J., Kolm, P. N., Pachamanova, D. A., & Focardi, S. M. (2007). Robust Portfolio Optimization. *The Journal of Portfolio Management*, 33(3), 40-48.
6. Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3), 595-612.
7. Pflug, G., & Wozabal, D. (2007). Ambiguity in portfolio selection. *Quantitative Finance*, 7(4), 435-442.
8. **Mohajerin Esfahani, P., & Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171(1), 115-166.**
9. **Blanchet, J., Chen, L., & Zhou, X. Y. (2022). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management science*, 68(9), 6382-6410.**
10. Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2, 21-42.
11. Blanchet, J., Kang, Y., Murthy, K., & Zhang, F. (2019). Data-driven optimal transport cost selection for distributionally robust optimization. In 2019 winter simulation conference (WSC) (pp. 3740-3751). IEEE.
12. Costa, G., & Iyengar, G. N. (2023). Distributionally robust end-to-end portfolio construction. *Quantitative Finance*, 23(10), 1465-1482.

Appendix. A

Max 문제로 전환

$$- \min_{Q \in \mathcal{U}_\delta(\hat{P}_N)} \mathbb{E}_Q[w^\top \xi] = -\max_Q \mathbb{E}_Q[(-w)^\top \xi]$$

Strong duality 적용 (Esfahani & Kuhn (2018) - step 3)

$$- \max_Q \mathbb{E}_Q[(-w)^\top \xi] = \inf_{\lambda \geq 0} \left\{ \lambda \delta + \frac{1}{N} \sum_{i=1}^N \Phi_\lambda(\xi_i) \right\}, \Phi_\lambda(x_i) := \sup_{\xi} \{(-w)^\top \xi - \lambda \|\xi - x_i\|_q^2\}$$

$\Phi_\lambda(x_i)$ 계산

$$- \Phi_\lambda(x_i) = \sup_{\Delta} \{(-w)^\top \Delta - \lambda \|\Delta\|_q^2\} - w^\top x_i, \Delta := \xi - x_i$$

$$- \sup_{\Delta} \{(-w)^\top \Delta - \lambda \|\Delta\|_q^2\} = \sup_{t \geq 0} \sup_{\|\Delta\|_q = t} \{(-w)^\top \Delta - \lambda t^2\} = \sup_{t \geq 0} \{\|w\|_p t - \lambda t^2\} = \frac{\|w\|_p^2}{4\lambda} \quad (t^* = \frac{\|w\|_p}{2\lambda})$$

$$- \Phi_\lambda(x_i) = \frac{\|w\|_p^2}{4\lambda} - w^\top x_i$$

산술-기하평균 - λ 최적화

$$- \max_Q \mathbb{E}_Q[(-w)^\top \xi] = \inf_{\lambda \geq 0} \left\{ \lambda \delta + \frac{\|w\|_p^2}{4\lambda} \right\} - \frac{1}{N} \sum_{i=1}^N w^\top x_i = \sqrt{\delta} \|w\|_p - w^\top \bar{x}$$

$$- \text{AM-GM: } \inf_{\lambda} \left\{ \lambda a + \frac{b}{\lambda} \right\} = 2\sqrt{ab},$$

α 고정 - 목적함수 완화

$$- \text{Let } \mathbb{E}_Q[w^\top \xi] = \alpha, \text{ Var}_Q(w^\top \xi) = w^\top \mathbb{E}_Q[\xi \xi^\top] w - \alpha^2,$$

$$- \max_{Q \in \mathcal{U}_\delta(\hat{P}_N)} w^\top \text{Var}_Q(\xi) w = \max_{\alpha \geq \bar{\alpha}} \left\{ \max_{Q, \mathbb{E}_Q[w^\top \xi] = \alpha} w^\top \mathbb{E}_Q[\xi \xi^\top] w - \alpha^2 \right\}$$

Appendix. B

제약이 2개인 문제

$$\begin{aligned}
 - \max_{Q, \mathbb{E}_Q[w^\top \xi] = \alpha} w^\top \mathbb{E}_Q[\xi \xi^\top] w &= \inf_{\lambda_1 \geq 0, \lambda_2} \sup_Q \{ \mathbb{E}_Q[(w^\top \xi)^2] - \lambda_1 c(x, \xi) - \lambda_2 w^\top \xi \} + \lambda_1 \delta + \lambda_2 \alpha \\
 &= \inf_{\lambda_1 \geq 0, \lambda_2} \frac{1}{N} \sum \sup_{\xi} \{ (w^\top \xi)^2 - \lambda_1 c(x_i, \xi) - \lambda_2 w^\top \xi \} + \lambda_1 \delta + \lambda_2 \alpha; \Phi(x_i) := \sup_{\xi} \{ (w^\top \xi)^2 - \lambda_1 c(x_i, \xi) - \lambda_2 w^\top \xi \}
 \end{aligned}$$

$\Phi(x_i)$ 계산

$$\begin{aligned}
 - \Phi(x_i) &= \sup_{\xi} \{ (w^\top \xi)^2 - \lambda_1 c(x_i, \xi) - \lambda_2 w^\top \xi \} = \sup_{\xi} \{ (w^\top \xi)^2 - \lambda_1 \|\xi - x_i\|_q^2 - \lambda_2 w^\top \xi \} \\
 &= \sup_{\Delta} \{ (w^\top (x_i + \Delta))^2 - \lambda_1 \|\Delta\|_q^2 - \lambda_2 w^\top (x_i + \Delta) \}; \Delta := \xi - x_i \\
 &= \sup_{\Delta} \{ (w^\top x_i)^2 + 2(w^\top x_i)(w^\top \Delta) + (w^\top \Delta)^2 - \lambda_1 \|\Delta\|_q^2 - \lambda_2 w^\top x_i - \lambda_2 w^\top \Delta \} \\
 &= (w^\top x_i)^2 - \lambda_2 w^\top x_i + \sup_{\Delta} \{ (w^\top \Delta)^2 - \lambda_1 \|\Delta\|_q^2 + (2w^\top x_i - \lambda_2)(w^\top \Delta) \} \\
 - \sup_{\Delta} \{ (w^\top \Delta)^2 - \lambda_1 \|\Delta\|_q^2 + (2w^\top x_i - \lambda_2)(w^\top \Delta) \} &= \sup_{t \geq 0} \sup_{s \in [-t\|w\|_p, t\|w\|_p]} \{ s^2 - \lambda_1 t^2 + (2w^\top x_i - \lambda_2)s \}; s = w^\top \Delta \\
 &= \sup_{t \geq 0} \{ (\|w\|_p^2 - \lambda_1)t^2 + |2w^\top x_i - \lambda_2| \cdot \|w\|_p t \} \\
 - \Phi(x_i) &= (w^\top x_i)^2 - \lambda_2 w^\top x_i + \sup_{t \geq 0} \{ (\|w\|_p^2 - \lambda_1)t^2 + |2w^\top x_i - \lambda_2| \cdot \|w\|_p t \} \\
 &= (w^\top x_i)^2 - \lambda_2 w^\top x_i + \frac{(2w^\top x_i - \lambda_2)^2 \|w\|_p^2}{(4\lambda_1 - \|w\|_p^2)}, \lambda_1 > \|w\|_p^2
 \end{aligned}$$

Appendix. B (conti.)

λ 최적화

- $H(\lambda_1, \lambda_2) := \frac{1}{N} \sum_i \Phi(x_i) + \lambda_1 \delta + \lambda_2 \alpha$
- λ_2 에 대해 미분 후 정리 $\rightarrow \lambda_2^* = 2\alpha - \frac{2C\lambda_1}{\|w\|_p^2}; C = \alpha - w^\top \bar{x}$
- $\Phi(x_i)|_{\lambda_2^*} = (w^\top x_i)^2 - \left(2\alpha - \frac{2C\lambda_1}{\|w\|_p^2}\right) w^\top x_i + \frac{\left(2w^\top x_i - 2\alpha - \frac{2C\lambda_1}{\|w\|_p^2}\right)^2 \|w\|_p^2}{(4\lambda_1 - \|w\|_p^2)}, \quad C := \alpha - w^\top \bar{x}, d_i := w^\top (x_i - \bar{x}), \kappa := \lambda_1 - \|w\|_p^2 > 0$

$$= d_i^2 \left(1 + \frac{\|w\|_p^2}{\kappa}\right) + d_i(2w^\top \bar{x} - \lambda_2^* + 2C) + (w^\top \bar{x})^2 - \lambda_2^* w^\top \bar{x} + \frac{C^2 \kappa}{\|w\|_p^2}$$
- $\frac{1}{N} \sum_i \Phi(x_i) = \left(1 + \frac{\|w\|_p^2}{\kappa}\right) w^\top \text{Var}(x) w + (w^\top \bar{x})^2 - \lambda_2^* w^\top \bar{x} + \frac{C^2 \kappa}{\|w\|_p^2}$

$H(*)$ 정리 - 산술-기하평균

- $H(\kappa) = \frac{1}{N} \sum_i \Phi(x_i) + (\kappa + \|w\|_p^2) \delta + \lambda_2^* \alpha = w^\top \text{Var}(x) w + (w^\top \bar{x})^2 + 2\alpha C - 2C^2 + \delta \|w\|_p^2 + \kappa \left(\delta - \frac{C^2}{\|w\|_p^2}\right) + \frac{\|w\|_p^2 w^\top \text{Var}(x) w}{\kappa}$

$$= w^\top \text{Var}(x) w + (w^\top \bar{x})^2 + 2\alpha C - 2C^2 + \delta \|w\|_p^2 + 2\sqrt{\delta \|w\|_p^2 - C^2} \sqrt{w^\top \text{Var}(x) w}$$
- $h(\alpha, w) = \mathbb{E}_Q[(w^\top \xi)^2] + 2(\alpha - w^\top \bar{x}) w^\top \bar{x} + \delta \|w\|_p^2 + 2\sqrt{\delta \|w\|_p^2 - (\alpha - w^\top \bar{x})^2} \sqrt{w^\top \text{Var}(x) w} = \left(\sqrt{w^\top \text{Var}(x) w} + \sqrt{\delta} \|w\|_p\right)^2 + \alpha^2$

Appendix. C

$$\mathcal{U}_\delta(P_{N,t}; M_t) = \left\{ Q : \inf_{\pi \in \Pi(P_{N,t}, Q)} \mathbb{E}_\pi [(X - \Xi)^\top M_t (X - \Xi)] \leq \delta \right\}, \quad M_t \succ 0$$

$$\eta = \sqrt{\delta}, \quad q_{M_t}(w) = \sqrt{w^\top M_t^{-1} w}.$$

원래 DRMV primal:

$$\begin{aligned} \min_w \quad & \max_{Q \in \mathcal{U}_\delta} \text{Var}_Q(w^\top \xi) \\ \text{s.t.} \quad & \min_{Q \in \mathcal{U}_\delta} \mathbb{E}_Q[w^\top \xi] \geq \bar{\rho}_t, \\ & \mathbf{1}^\top w = 1, \quad w \geq 0. \end{aligned}$$

$M_t \succ 0$ 이므로

$$\tilde{x} = M_t^{1/2} x, \quad \tilde{\xi} = M_t^{1/2} \xi, \quad \tilde{w} = M_t^{-1/2} w$$

$$(x - \xi)^\top M_t (x - \xi) = \|\tilde{x} - \tilde{\xi}\|_2^2, \quad \|\tilde{w}\|_2 = q_{M_t}(w).$$

따라서 Mahalanobis 문제를 Euclidean BCZ 문제로 환원.

Worst-case mean

$$\Delta = \Xi - X, \quad \mathbb{E}_\pi[\Delta^\top M_t \Delta] \leq \eta^2$$

$$\begin{aligned} \mathbb{E}_Q[w^\top \xi] &= w^\top \bar{x}_t + \mathbb{E}_\pi[w^\top \Delta] \\ &\geq w^\top \bar{x}_t - \sqrt{w^\top M_t^{-1} w} \sqrt{\mathbb{E}_\pi[\Delta^\top M_t \Delta]} \\ &\geq w^\top \bar{x}_t - \eta q_{M_t}(w). \end{aligned}$$

Equality-achieving perturbation:

$$\Delta_\mu^* = -\eta \frac{M_t^{-1} w}{q_{M_t}(w)}$$

$$(\Delta_\mu^*)^\top M_t \Delta_\mu^* = \eta^2, \quad w^\top \Delta_\mu^* = -\eta q_{M_t}(w).$$

따라서

$$\min_{Q \in \mathcal{U}_\delta} \mathbb{E}_Q[w^\top \xi] = w^\top \bar{x}_t - \eta q_{M_t}(w).$$

Worst-case variance

$$\begin{aligned} \sqrt{\text{Var}_Q(w^\top \xi)} &= \sqrt{\text{Var}(w^\top X + w^\top \Delta)} \\ &\leq \sqrt{w^\top \widehat{\Sigma}_t w} + \sqrt{\mathbb{E}[(w^\top \Delta)^2]} \\ &\leq \sqrt{w^\top \widehat{\Sigma}_t w} + \eta q_{M_t}(w). \end{aligned}$$

$$s = \sqrt{w^\top \widehat{\Sigma}_t w}, \quad Z = \frac{w^\top X - w^\top \bar{x}_t}{s}$$

Equality-achieving perturbation:

$$\Delta_V^* = \eta \frac{M_t^{-1} w}{q_{M_t}(w)} Z$$

$$\mathbb{E}[(\Delta_V^*)^\top M_t \Delta_V^*] = \eta^2, \quad w^\top (\Xi - \mathbb{E}\Xi) = (s + \eta q_{M_t}(w)) Z.$$

따라서

$$\max_{Q \in \mathcal{U}_\delta} \text{Var}_Q(w^\top \xi) = \left(\sqrt{w^\top \widehat{\Sigma}_t w} + \eta q_{M_t}(w) \right)^2.$$

Tractable reformulation

$$\begin{aligned} \min_w \quad & \sqrt{w^\top \widehat{\Sigma}_t w} + \eta q_{M_t}(w) \\ \text{s.t.} \quad & w^\top \bar{x}_t - \eta q_{M_t}(w) \geq \bar{\rho}_t, \\ & \mathbf{1}^\top w = 1, \quad w \geq 0. \end{aligned}$$

$$C_t^\top C_t = \widehat{\Sigma}_t, \quad L_t^\top L_t = M_t^{-1}$$

$$\begin{aligned} \min_{w, u, v} \quad & u + \eta v \\ \text{s.t.} \quad & \|C_t w\|_2 \leq u, \\ & \|L_t w\|_2 \leq v, \\ & w^\top \bar{x}_t - \eta v \geq \bar{\rho}_t, \\ & \mathbf{1}^\top w = 1, \quad w \geq 0. \end{aligned}$$