

Structure-Aware Distributionally Robust Portfolio via Inverse Reinforcement Learning

[paper review & proposal]

진영봉

FRE LAB

Jan 27, 2026

Background

- The Dilemma of Portfolio Optimization

- MVO (Mean-Variance)

Formulation: $\min_w \frac{\lambda}{2} w^T \hat{\Sigma} w - w^T \hat{\mu}$

Limit: 입력변수의 추정 오차에 민감하며, 이로 인해 out-of-sample(OOS) 성능이 낮음 [1]

- RO (Robust Optimization): 파라미터 공간의 결정론적 최악 상황을 가정

Formulation : $\min_w \max_{(\mu, \Sigma) \in \mathcal{U}} \left\{ \frac{\lambda}{2} w^T \Sigma w - w^T \mu \right\}$

불확실성 집합의 형태

- Box Uncertainty: $\mathcal{U}_{box} = \{(\mu, \Sigma): \mu_i \in [\underline{\mu}_i, \overline{\mu}_i], \Sigma \succcurlyeq 0\}$; **장점:** 계산이 쉽고, 각 자산별 구간 설정 가능 **단점:** 각 자산간 상관관계 고려 X, 너무 보수적일 가능성 [2,3]

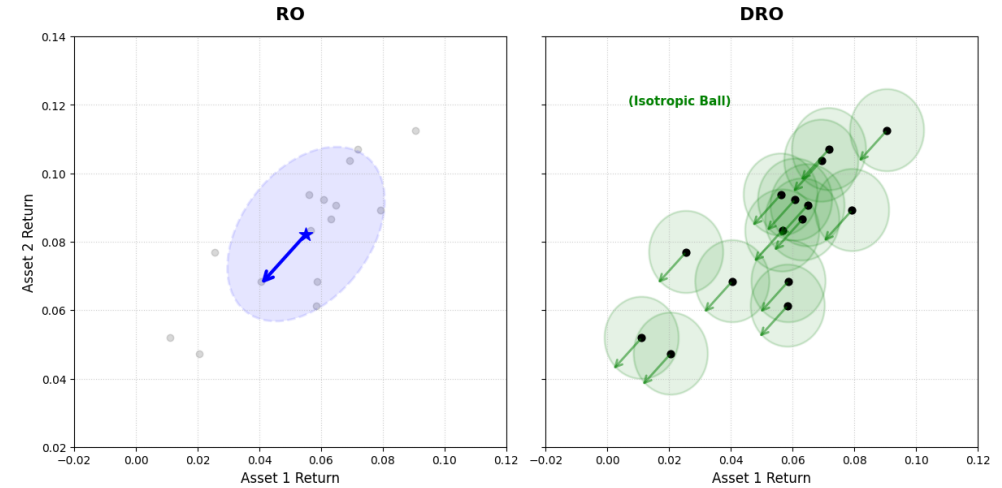
- Ellipsoidal Uncertainty: $\mathcal{U}_{ellip} = \{\mu: \|\mu - \hat{\mu}\|_{\Sigma^{-1}} \leq \delta\}$; **장점:** 공분산 구조를 반영, closed solution **단점:** δ 를 정하는데 구체적 근거 없음 [3]

Limit: 확률론적 구조를 무시하여 모든 불확실성이 동등하게 취급되며, 이로 인해 구조적 정보 활용 제한 → 과도한 보수성

- DRO (Distributionally Robust Optimization): 확률 측도 공간에서의 최악의 상황 가정

:Formulation : $\min_w \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} [L(w, \xi)]$

: Limit: 모호성 집합(Ambiguity Set)의 형태를 어떻게 정의할 것인가?



Theoretical Foundation of DRO - Definition

- **General Primal Problem:** $\min_w \sup_{\mathbb{p} \in \mathcal{P}} \mathbb{E}_{\mathbb{p}}[L(w, \xi)]$
 - $L(w, \xi)$: 포트폴리오 비중 w 과 수익률 ξ (ambiguity set) 이 주어졌을 때 발생하는 손실
 - $\mathbb{E}_{\mathbb{p}}[\cdot]$: 특정 확률 분포 \mathbb{p} 하에서의 기대 손실 \rightarrow 시장이 \mathbb{p} 라는 분포(혹은 시나리오)대로 움직인다면 평균적으로 얼마나 손해를 보는가?
 - $\sup_{\mathbb{p} \in \mathcal{P}}$: 정의된 모호성 집합 \mathcal{P} 안에 있는 수 많은 분포들 중, 기대 손실을 가장 크게 만드는 분포를 찾음 (최악의 시나리오 선택)
 - \min_w : 최악의 시나리오상에서 손실을 최소화하는 방어적인 비중 w 를 선택
- **Components:**
 - Ξ : 불확실한 파라미터 (수익률)의 support set
 - $\mathcal{M}(\Xi)$: Ξ 위에서 정의된 모든 확률 측도 공간
 - $\mathcal{P} \subseteq \mathcal{M}(\Xi)$: ambiguity set. 참 분포 \mathbb{p}_{true} 를 포함한다고 신뢰할 수 있는 분포들의 집합
 - : 미래의 진짜 수익률 분포 \mathbb{p}_{true} 은 알 수 없지만, 적어도 과거 데이터로 만든 경험적 분포 $\hat{\mathbb{p}}_{true}$ 와 크게 다르지 않을 것이라고 가정
 - : $\hat{\mathbb{p}}_{true}$ 주변의 특정 거리 ϵ 안에 있는 분포들 중 정답이 있다고 가정하고 집합을 구성
 - : 이 집합의 크기를 너무 작게 잡으면 MVO(데이터 과신)가 되고, 너무 크게 잡으면 RO(지나친 보수성)가 됨

Theoretical Foundation of DRO - Constructing Ambiguity Sets

- **Key Challenge:** ambiguity set(\mathcal{P})를 어떻게 정의하는지에 따라 (i) out-of-sample 성과와 (ii) 계산 복잡도가 결정됨
- **Moment-based Approach:**
 - 가장 직관적이고 클래식한 방법으로, 분포의 구체적인 모양보다는 통계적 요약치를 활용
 - 핵심 아이디어: 진짜 분포가 어떻게 생겼는지는 모르겠지만, support / mean / covariance 같은 요약치는 신뢰할 수 있다고 가정
- **Distance/Divergence-based Approach:**
 - 최신 DRO연구의 주류를 이루는 방식으로, 분포의 형태가 기준 분포와 얼마나 가까운지를 고려함
 - 핵심 아이디어: 진짜 분포는 내가 가진 데이터 분포와 모양이 비슷할 것이기 때문에 모양이 조금 다른 분포들 까지만 고려
 - A. f -Divergence (KL, χ^2 등)
 - 두 확률 분포의 PDF를 직접적으로 비교(density / likelihood ratio)하고, 이를 기반으로 divergence ball (ambiguity set) 구성
 - B. Wasserstein Distance (Current Standard)
 - Wasserstein 거리를 측정하여 두 분포 사이에서 거리 값을 산출하여 유사성을 측정

Literature Review: Moment-based DRO [4]

- **Definition:** \mathbb{P}_{true} 은 모르지만 $\hat{\mu}, \hat{\Sigma}$ 어느정도 믿을 수 있으며, 추정치 주변의 특정 신뢰구간 안에 있는 모든 분포를 고려

$$\mathcal{P}_{moment}(\gamma_1, \gamma_2) = \left\{ P \in \mathcal{M}(\Xi): \begin{array}{l} (\mathbb{E}_p[\xi] - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbb{E}_p[\xi] - \hat{\mu}) \leq \gamma_1 \\ \mathbb{E}_p[(\xi - \hat{\mu})(\xi - \hat{\mu})^T] \leq \gamma_2 \hat{\Sigma} \end{array} \right\}$$

- $(\mathbb{E}_p[\xi] - \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbb{E}_p[\xi] - \hat{\mu}) \leq \gamma_1$: 실제 평균($\mathbb{E}_p[\xi]$)이 추정 평균 ($\hat{\mu}$)을 중심으로 하는 타원안에 있음
- $\mathbb{E}_p[(\xi - \hat{\mu})(\xi - \hat{\mu})^T] \leq \gamma_2 \hat{\Sigma}$: 실제 2차 모멘트가 추정 공분산($\hat{\Sigma}$)의 일정 배수보다 작거나 같음. → 실제 리스크가 관측치보다 터무니 없이 크지는 않을 것이라는 가정

Results:

- S&P500에 포함된 30개 주요 종목; 4개의 자산군을 랜덤하게 뽑아 포트폴리오 구성 → 300회 반복
- **SP(Stochastic Programming)**: ambiguity set 없이 과거 추정치가 정답이라고 가정 (MVO와 동일)
- **Popescu's DRPO**: 초기 형태의 Moment-based DRO (1,2차 모멘트를 공유하는 다양한 분포 활용)
- **DRPO (proposed)**: 제안된 모델, 1,2차 모멘트의 불확실성까지 고려
- 불확실한 시장 상황(닷컴버블, 9/11)에서도 제안한 모델이 robust한 성능 달성

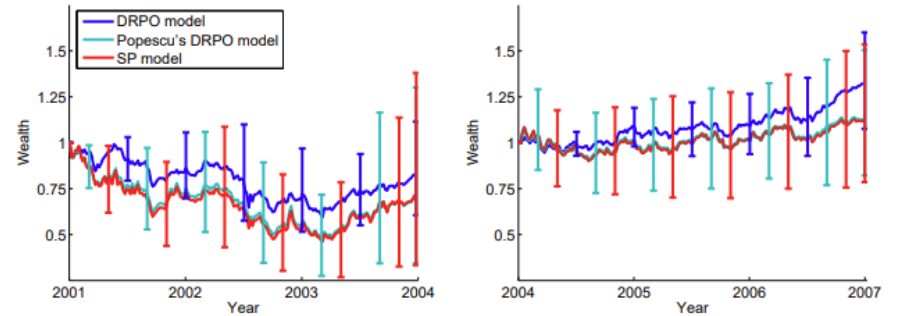


Figure 1 Comparison of wealth evolution in 300 experiments conducted over the years 2001-2007 using three different portfolio optimization models. For each model, the figures indicate periodically the 10% and 90% percentile of the wealth distribution in the set of experiments.

pros and cons:

- 장점: 계산 효율성이 뛰어남(Tractability). 다항시간 내에 SDP로 풀 수 있음을 증명
- 한계: 1,2차 모멘트만 활용하기 때문에 Skewness나 Kurtosis같은 고차 모멘트 정보를 포착하지 못함

Literature Review: Distance/Divergence-based DRO [5]

- **Definition:** \mathbb{P}_{true} 은 모르지만, 관측된 데이터의 경험적 분포 $\hat{\mathbb{P}}_N$ 과 Wasserstein Distance가 일정 수준 이내인 모든 확률 분포를 고려

$$\mathbb{B}_\varepsilon(\hat{\mathbb{P}}_N) := \{\mathbb{Q} \in \mathcal{M}(\mathcal{E}) : d_w(\hat{\mathbb{P}}_N, \mathbb{Q}) \in \varepsilon\}, d_w(\mathbb{Q}_1, \mathbb{Q}_2) = \inf_{\Pi} \left\{ \int_{\mathcal{E}^2} \|\xi_1 - \xi_2\| \Pi(d\xi_1, d\xi_2) \right\}, \|\cdot\| \text{은 } l_p \text{ norm}$$

- $\hat{\mathbb{P}}_N := \frac{1}{N} \sum_{i=1}^N \delta_{\xi_i}$, δ_{ξ} 는 Dirac 분포로 확률 질량이 ξ 에만 집중되어 있음을 나타내며, 관측된 N 개의 점 하나하나가 똑같은 확률을 나누어 주었음을 의미
- 관측된 데이터 점들 주변으로 반경 ε 만큼 떨어진, 실현 가능한 모든 시나리오들의 집합
- RO, Moment-based DRO 방식 대비 support set 자체가 유동적

Results:

- 10개의 자산시장 모델 (실제 데이터 X): $\xi_i = \psi + \zeta_i$, $\psi \sim N(0, 2\%), \zeta_i \sim N(i \times 3\%, i \times 2.5\%) \rightarrow$ 각 자산의 실제 분포를 알고있는 실험 환경 (OOS 성능 파악)
- ε (x축)을 키울 수록 **실선**이 내려감 (성능 개선; MVO의 overfitting 문제 개선)
- ε 을 키울 수록 **점선**이 올라감 (신뢰도 상승; 허용된 리스크 상한을 초과하지 않을 확률)
- ε 이 너무 작으면 과적합, 너무 크면 과도하게 보수적
- 데이터 적을때 DRO 성능이 큰 폭으로 개선

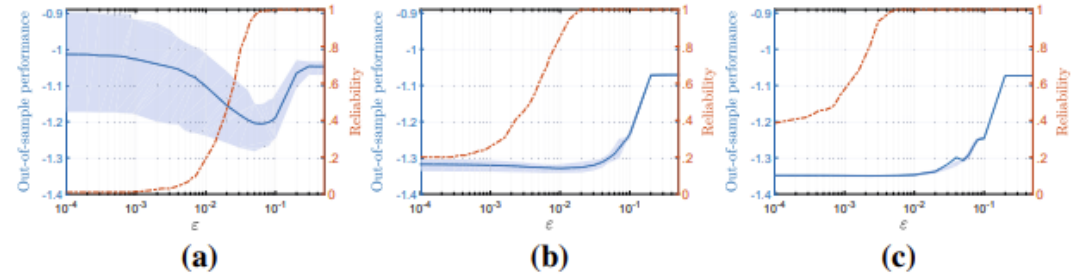


Fig. 5 Out-of-sample performance $J(\hat{x}_N(\varepsilon))$ (left axis, solid line and shaded area) and reliability $\mathbb{P}^N[J(\hat{x}_N(\varepsilon)) \leq \hat{J}_N(\varepsilon)]$ (right axis, dashed line) as a function of the Wasserstein radius ε and estimated on the basis of 200 simulations.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

pros and cons:

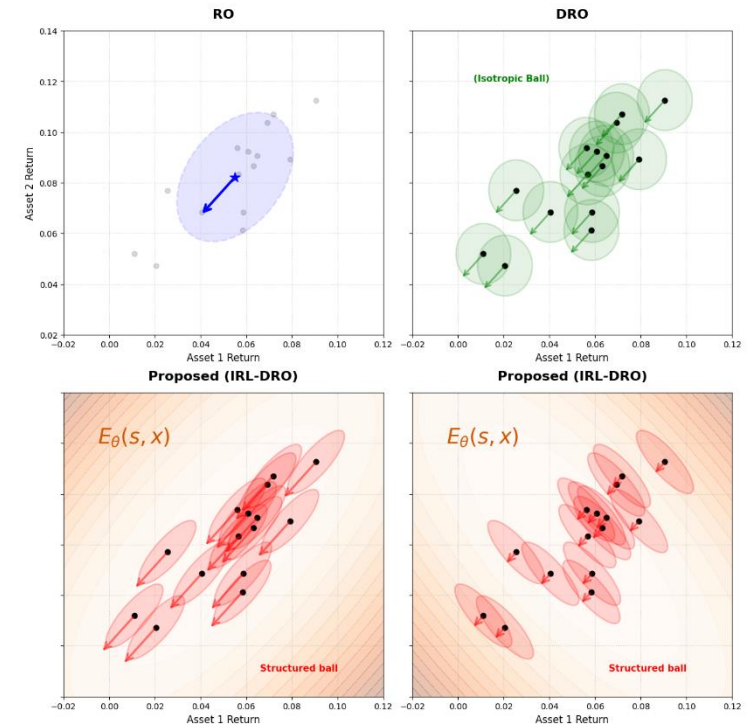
- 장점: 데이터가 부족한 상황에서도 과적합을 효율적으로 방지
무한차원에 해당하는 문제를 볼록 최적화 문제로 변환
- 한계: 이상적인 통제 환경에서 수행됨 (시뮬레이션 데이터를 활용) - 실제 데이터를 활용한 후속연구 존재 [6,7,8] // 최근 DL/RL을 활용한 DRO 연구도 수행됨 [9,10,11]

Comparative Analysis

- Research gap

- 기존 Wasserstein 거리 기반 DRO는 불확실성 집합이 모든 방향으로 균일한 Isotropic 형태를 갖도록 설계됨
- 금융 자산간 상관관계는 비대칭/비등방성을 띄며[12], 위기 시 특정 방향으로 위험이 증폭 되지만, Isotropic-sphere 형태의 방어막은 시장의 이러한 구조적 위험을 반영하지 않고 있음

Feature	MVO	RO	DRO	SA-DRO (proposed)
Uncertainty / Ambiguity set	Point Estimate ($\hat{\mu}, \hat{\Sigma}$)	Fixed set (Ellipsoid/Box)	Wasserstein Ball	Structure-Aware Ellipsoid
Geometry / Metric	Sample Statistics	Worst-case Bounds	Euclidean Distance (l_p family)	Learned Mahalanobis Metric (via IRL)
Ambiguity Shape	N/A (point)	Static (Pre-defined)	Isotropic	Anisotropic (Adaptive Ellipsoid)
Adaptivity	None	Parameter worst-case	Distributional worst-case	State-Dependent Geometry
Data Usage	Parameter Estimation	Bound Estimation	Empirical Distribution	Generative Process Modeling (Energy/Reward)
Robustness	None (Sensitive)	High (Conservative)	High (Data-driven)	High (Structure-Aware)



Methodology

- **Research goal:** 시장의 내재적 정보를 학습하여 불확실성 집합 구조를 반영하는 SA-DRO (Structure-Aware DRO)를 제안
시장 데이터 자체를 시장 참여자들의 집단 지성이 반영된 Expert Demonstration으로 간주하고 IRL을 통해 시장의 내재적 구조를 학습

$$\min_w \sup_{Q \in \mathcal{U}_\epsilon(P_n)} \mathbb{E}_Q[-w^\top \xi]$$

$$\begin{aligned} \mathcal{U}_\epsilon(P_n) &= \{Q: W(Q, P_n) \leq \epsilon\} \\ W(Q, P_n) &= \inf_{\pi \in \Pi(Q, P_n)} \mathbb{E}_\pi[\|\xi - \xi'\|] \end{aligned}$$

$$\mathbb{B}_\epsilon(P_n) := \{Q \in \mathcal{M}(\Xi) : W(Q, P_n) \leq \epsilon\}$$



$$\min_w \sup_{Q \in \mathcal{U}_\epsilon^{SA}(P_n, s)} \mathbb{E}_Q[-w^\top \xi]$$

$$\begin{aligned} \mathcal{U}_\epsilon^{SA}(P_n, s) &= \{Q: d_{M(s)}(Q, P_n) \leq \epsilon\} \\ d_{M(s)}(Q, P_n) &= \inf_{\pi \in \Pi(Q, P_n)} \mathbb{E}_\pi[\sqrt{(\xi - \xi')^\top M(s)(\xi - \xi')}] \\ M(s) &:= H_\theta(s) + \rho I, \rho > 0 \end{aligned}$$

$$\mathbb{B}_\epsilon^{SA}(P_n) := \{Q \in \mathcal{M}(\Xi) : d_{M(s)}(Q, P_n) \leq \epsilon\}$$

- $P_n := \frac{1}{N} \sum_{i=1}^N \delta_{\hat{\xi}_i}, \delta_{\hat{\xi}_i}(\xi) = \begin{cases} 1 & \text{if } \xi = \hat{\xi}_i \\ 0 & \text{if } \xi \neq \hat{\xi}_i \end{cases}$
- $H_\theta(s) = \nabla_\xi^2 E_\theta(s, \xi)|_{\xi=\bar{\xi}(s)}, E_\theta: S \times \Xi \rightarrow \mathbb{R}$

Methodology

- Notation

- $s_t \in S$: 시장 상태 (예: 과거 k일간의 수익률, 거시경제 지표 등)
- $r_t \in \mathbb{R}^N$: 실제 관측된 수익률 (realized return, expert action)
- $\xi \in \mathbb{R}^N$: 확률 변수로서의 수익률 (Random Variable) $\bar{\xi}$: Generator가 만들어내는 수익률, ξ : ambiguity variable (generator와 무관)

- IRL의 적용: Market as an Expert

- 시장이 항상 효율적이고 완벽한 것은 아니지만, 시장 데이터 자체를 전문가의 행동으로 간주 (expert demonstration)
- 하지만, 완벽하지 않은 전문가 데이터라도 Feasible Reward Set의 범위를 좁히는 데 기여할 수 있음[13]
- 즉, 시장은 어떤 선호/위험 구조에 의해 "최적의 분포(P^*)에서 수익률(r_t)"을 발생시키는 시스템으로 놓고 IRL 학습을 통해 ambiguity set을 추정하고자 함

- Energy Function Learning via AIRL:

- Discriminator $D_\theta(s, \xi)$ / Energy network: $E_\theta(s, \xi) = -f_\theta((s, \xi))$
- Generator (Policy): $\pi_\phi(\xi|s)$ (조건부로 ξ 샘플링)
- Generator는 최종 산출물이 아니라, Discriminator가 trivial solution에 빠지지 않도록 "negative samples"를 제공하는 학습 파트너

$$D_\theta(s, \xi) = \frac{\exp(f_\theta(s, \xi))}{\exp(f_\theta(s, \xi)) + \pi_\phi(\xi|s)}$$

- $L_D(\theta) = -\mathbb{E}_{r \sim \text{Data}}[\log D_\theta(s, r)] - \mathbb{E}_{\bar{\xi} \sim \pi_\phi}[\log(1 - D_\theta(s, \bar{\xi}))]$: E_θ 는 "실제 시장 데이터가 자주 위치 하는 곳"의 에너지를 낮추는 형태로 학습됨
- $L_G(\phi) = \mathbb{E}_{\bar{\xi} \sim \pi_\phi}[\log(1 - D_\theta(s, \bar{\xi})) - \log D_\theta(s, \bar{\xi})] \approx \mathbb{E}_{\bar{\xi} \sim \pi_\phi}[-E_\theta(s, \bar{\xi})]$: D_θ 가 진짜 같다고 착각하게 만드는 (에너지가 낮은) 샘플을 생성하도록 학습

Methodology

- Metric induction (Energy to Geometry)

- 샘플 x 에 대해 판별자 $D(x)$ 가 최적화 되면, 이는 에너지 함수를 학습하는 것과 같고, 그 결과 $p(x) \propto \exp(-E(x))$ 형태의 분포를 따른다는 것이 증명됨 [14]
- 확률 공간에서의 효율적인 탐색을 위해 국소 곡률 정보를 활용됨 [15]

$$M(s) \triangleq \nabla_{\xi}^2 E_{\theta}(s, \xi) \Big|_{\xi=r_t} + \rho I$$

- 에너지 함수의 Hessian matrix상에서 실제 수익률 r_t 로 평가된 값을 사용
- 특정 자산 간의 수익률 변화에 따른 에너지 민감도(비선형 상관구조)를 포착
- $M(s)$ 는 현재 시장 상태에서 자산들의 움직임이 자연스러운지 아니면 구조적으로 거부되는지를 측정할 수 있도록 세팅
 - 금융위기 시, 모든 자산이 다 같이 폭락하는 현상 -> 에너지가 완만하게 변하고 $M(s)$ 가 작음 (자연스러운 시장 움직임)
 - 삼성전자가 10%하락했는데, SK하이닉스가 30% 상승한다 -> 시장 구조상 거의 불가능. 에너지가 급격하게 솟구치고 $M(s)$ 가 큼

- Anisotropic Ambiguity Set:

$$\mathbb{B}_{\epsilon}^{SA} := \left\{ Q : \inf_{\Pi} \int_{\mathbb{E}^2} \sqrt{(\xi - r)^T M(s) (\xi - r)} d_{\pi}(\xi, r) \leq \epsilon \right\}$$

- 모호성 집합의 중심은 generator ($\bar{\xi}$)가 아닌 실제 데이터 (r)
- 각 데이터 포인트(r_i)를 중심으로 에너지 지형 곡률 $M(s)$ 에 따라 찌그러진 타원체로 불확실성 방어할 수 있도록 설계
- 물리적 거리가 아닌, 시장 에너지 지형을 반영한 거리 사용

Methodology*

- Primal problem: ambiguity variable ξ 는 mahalanobis 거리 기준으로 ϵ 안에서 움직일 수 있음

$$\min_w \sup_Q \left\{ \mathbb{E}_Q[-w^\top \xi] : \int \|\xi - r\|_M d\pi \leq \epsilon \right\}, \|\xi - r_i\|_M = \sqrt{(\xi - r)^\top M (\xi - r)}$$

- Lagrangian:

$$\min_{w, \lambda \geq 0} \left(\lambda \epsilon + \sup_Q \mathbb{E}_Q[H(\xi)] \right), H(\xi) = -w^\top \xi - \lambda \|\xi - r\|_M$$

- $\sup_Q \mathbb{E}_Q[H(\xi)]$: 평균 점수가 최대가 되도록 분포(Q)를 조작 \rightarrow 가장 좋은 전략은 $H(\xi)$ 가 가장 높은 값을 갖는 ξ 에 확률을 몰아주는 것 (dirac 함수)
- (dirac 함수 조건하에서) $\sup_Q \mathbb{E}_Q[H(\xi)] = \mathbb{E}_{P_n} \left[\sup_{\xi} H(\xi) \right] \rightarrow$ 각 데이터 포인트 r_i 별 최악의 시나리오를 찾는 문제로 대체가능
- $\min_w \left(\lambda \epsilon + \frac{1}{N} \sum_{i=1}^N \sup_{\xi} (-w^\top \xi - \lambda \|\xi - r_i\|_M) \right)$

- Dual Problem:

$$\min_{w, \lambda, s} \left(\lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \right) \quad \rightarrow \quad \min_{w, \lambda, s} \left(\lambda \epsilon + \frac{1}{N} \sum_{i=1}^N s_i \right)$$

$$s_i \geq \sup_{\xi} (-w^\top \xi - \lambda \|\xi - r_i\|_M) \quad s_i \geq -w^\top r_i$$

$$\lambda \geq \|M^{-1/2} w\|_2$$

Results – Benchmark [7]

- Data

- 18종목의 ETF 데이터 사용 (weekly)
- 과거 5년치 데이터를 사용하여 파라미터 추정
 - 전체 구간 52*5 = 260 / training = 182, validation = 78
- 바로 다음주 포트폴리오 비중계산
- 수수료 고려 X

- Results

	1/N (B&H)	MVO	DRMV [7]
Ann. Return	9.33%	12.20%	12.91%
Sharpe ratio	0.7684	0.5751	0.7352
MDD	-25.42%	-35.24%	-29.95%

- Future Plan

- 베이스라인 모델(Wasserstein DRO)를 재현한 결과, 일반적인 MVO 보다 우수한 성과를 보여주는 것을 확인
- IRL 기반 메트릭 학습 및 적용 후, 추가적인 성과 개선이 있는지 확인
- 주식 외에 채권, 원자재 등 다양한 자산군으로 실험을 확장하고, 현실성을 위해 수수료 및 슬리피지 고려하여 실험

Figure 1: Out-of-Sample Performance Comparison

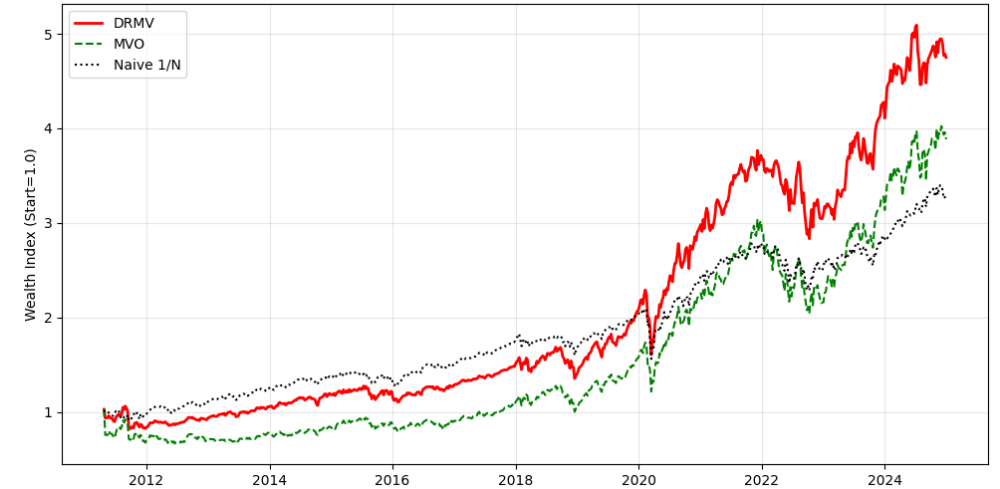
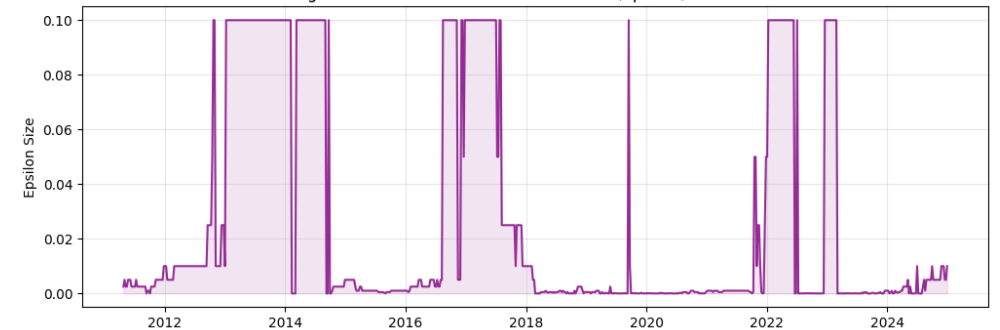


Figure 2: Calibrated Robustness Radius (Epsilon) Over Time



References

1. Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized' optimal?. *Financial analysts journal*, 45(1), 31-42.
2. Halldórsson, B. V., & Tütüncü, R. H. (2003). An interior-point method for a class of saddle-point problems. *Journal of Optimization Theory and Applications*, 116(3), 559-590.
3. Goldfarb, D., & Iyengar, G. (2003). Robust portfolio selection problems. *Mathematics of operations research*, 28(1), 1-38.
4. Delage, E., & Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3), 595-612.
5. 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.
6. Fonseca, D., & Junca, M. (2021). Wasserstein Distributionally Robust Optimization with Expected Value Constraints. *arXiv preprint arXiv:2111.04663*.
7. Blanchet, J., Chen, L., & Zhou, X. Y. (2022). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management science*, 68(9), 6382-6410.
8. Hosseini-Nodeh, Z., Khanjani-Shiraz, R., & Pardalos, P. M. (2023). Portfolio optimization using robust mean absolute deviation model: Wasserstein metric approach. *Finance Research Letters*, 54, 103735.
9. Costa, G., & Iyengar, G. N. (2023). Distributionally robust end-to-end portfolio construction. *Quantitative Finance*, 23(10), 1465-1482.
10. Lu, C. I., Sester, J., & Zhang, A. (2025). Distributionally Robust Deep Q-Learning. *arXiv preprint arXiv:2505.19058*.
11. Du, N., Liu, Y., & Liu, Y. (2020). A new data-driven distributionally robust portfolio optimization method based on wasserstein ambiguity set. *IEEE Access*, 9, 3174-3194.
12. Ang, A., & Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of financial Economics*, 63(3), 443-494.
13. Poiani, R., Gabriele, C., Metelli, A. M., & Restelli, M. (2024). Sub-optimal experts mitigate ambiguity in inverse reinforcement learning. *Advances in Neural Information Processing Systems*, 37, 85778-85823.
14. Finn, C., Christiano, P., Abbeel, P., & Levine, S. (2016). A connection between generative adversarial networks, inverse reinforcement learning, and energy-based models. *arXiv preprint arXiv:1611.03852*.
15. Girolami, M., & Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 73(2), 123-214.