PAKDD 2024 DMO-FinTech Workshop

Enhancing portfolio optimization with machine learning

Yongjae Lee

Department of Industrial Engineering
Artificial Intelligence Graduate School

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

Speaker

- **Yongjae Lee, Ph.D. (https://felab.unist.ac.kr)**
  - Associate Professor, UNIST IE & AIGS, 2024 to present
  - Assistant Professor, UNIST IE, 2018 to 2024
  - Ph.D. in Industrial and Systems Engineering, KAIST, 2016
  - B.S. in Computer Science and Mathematical Sciences, KAIST, 2011

  - Editorial board member, Journal of Financial Data Science
  - Guest editor, Special issue on statistical and machine learning for investor modelling, Journal of Behavioral Finance
  - Organizing committee, 5th International Conference on AI in Finance (ICAIF'24)

Section 0

Overview

Markowitz model

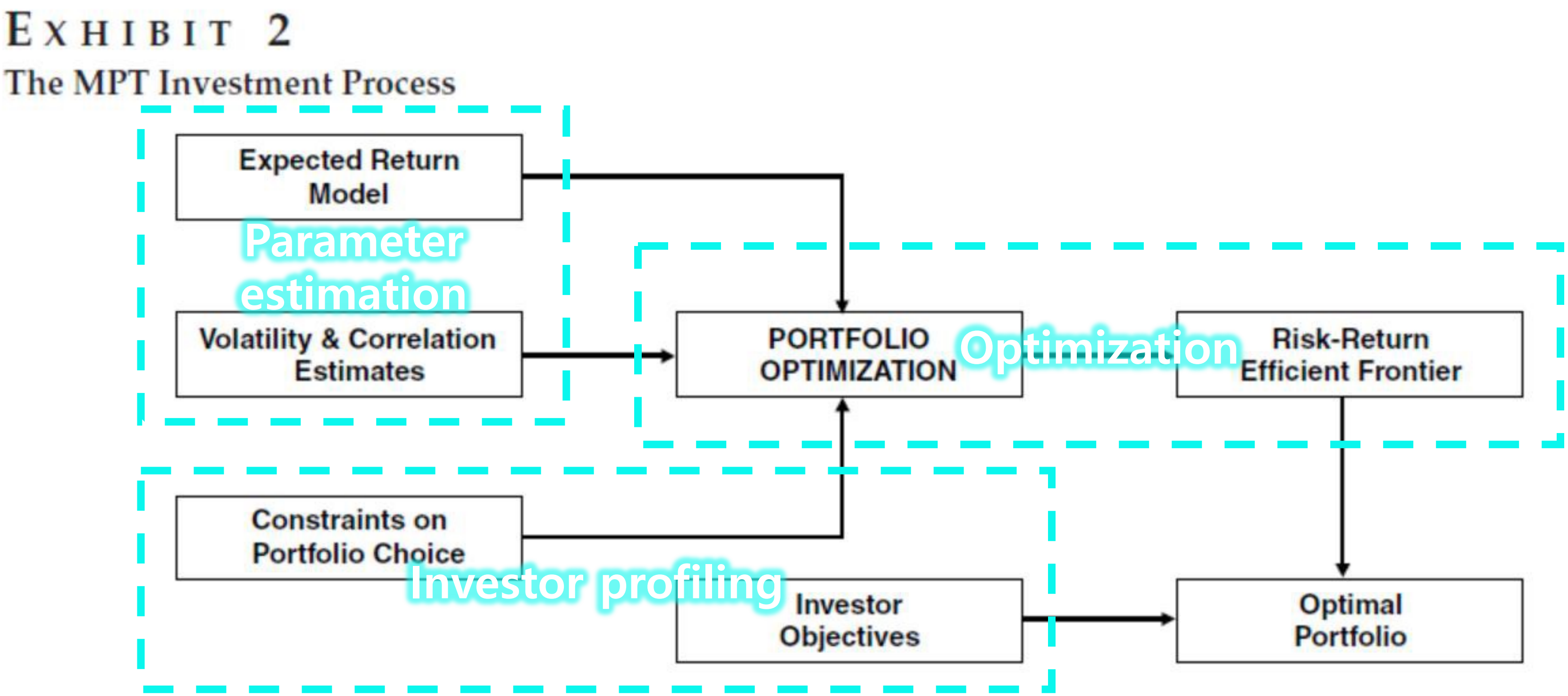
- **A quadratic programming (QP) problem**

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} w^T \Sigma w \\ \text{subject to} \quad & w^T \mu \geq \mu_0 \\ & \mathbf{1}^T w = 1 \end{aligned}$$

  - **Optimization variable (or decision variable)**
    - $w = [w_1, \ldots, w_n]^T \in \mathbb{R}^n$: portfolio weight vector
  - **Parameters**
    - $\mu = [\mu_1, \ldots, \mu_n]^T \in \mathbb{R}^n$: expected return vector
    - $\Sigma = [\sigma_{ij}]_{i,j=1}^{n} \in \mathbb{R}^{n \times n}$: return covariance matrix
    - $\mu_0 \in \mathbb{R}$: minimum required return
  - **Etc.**
    - $\mathbf{1} = [1,1, \ldots, 1]^T \in \mathbb{R}^n$: vector of ones
    - Short-selling is allowed

Mean-variance optimization process

- **Modern portfolio theory (MPT) investment process**
  (Fabozzi, Gupta, & Markowitz, 2002)



Fabozzi, F. J., Gupta, F., & Markowitz, H. M. (2002).
The legacy of modern portfolio theory. *The Journal of Investing*, *11*(3), 7-22.

Importance of parameter estimation

- **Optimal portfolio weight is sensitive to the input values**
  - Michaud (1989) even refers to mean-variance model as "estimation-error maximizers"
  - Broadie (1993) introduced the concept of true frontier, estimated frontier, and actual frontier
    (above figure is from Ceria & Stubbs (2005))
    - True frontier: efficient frontiers computed using the true expected returns (which is unobservable in advance)
    - Estimated frontier: efficient frontiers computed using the estimated expected returns
    - Actual frontier: returns and variances of estimated frontier portfolios computed using true (actual) returns
  - Chopra & Ziemba (1993)
    - They argued that, roughly speaking, errors in the expected returns are about 10 times more important than errors in the covariance matrix,
    - and errors in the variances are about twice as important as errors in the covariances

Michaud, R. O. (1989). The Markowitz optimization enigma: Is 'optimized'optimal?. *Financial Analysts Journal*, 45(1), 31-42.
Broadie, M. (1993). Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, *45*, 21-58.
Chopra, V. K., & Ziemba, W. T. (1993). The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2), 6.

Importance of parameter estimation

- **Optimal portfolio weight is sensitive to the input values**
  - Chung et al. (2022)
    - Suggested that correlation might be more important than means

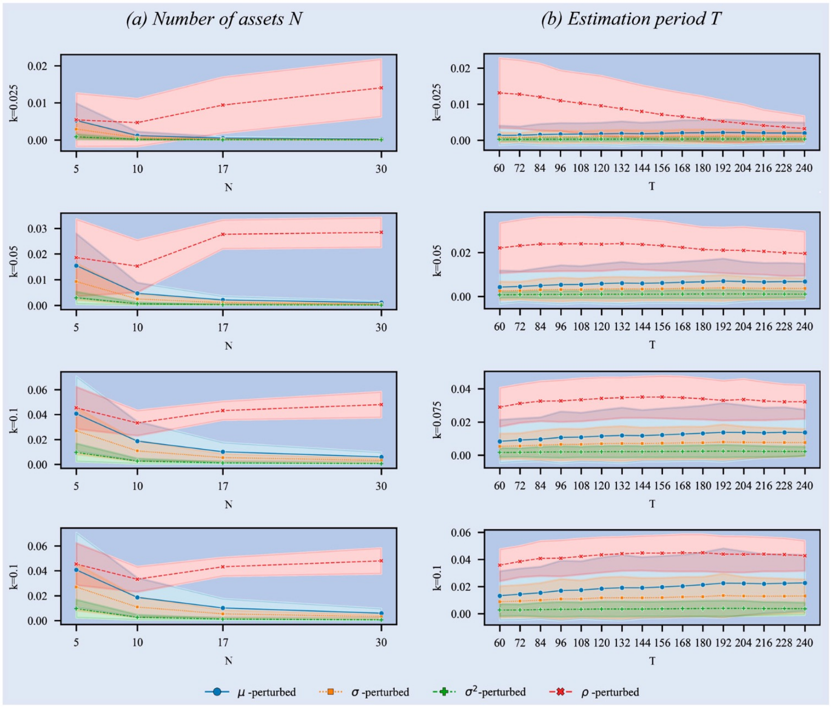


Figure 5. The effect of errors in MV parameters with (a) different number of assets and (b) different estimation periods.

Chung, M., Lee, Y., Kim, J. H., Kim, W. C., & Fabozzi, F. J. (2022).
The effects of errors in means, variances, and correlations on the mean-variance framework. *Quantitative Finance*, *22*(10), 1893-1903.

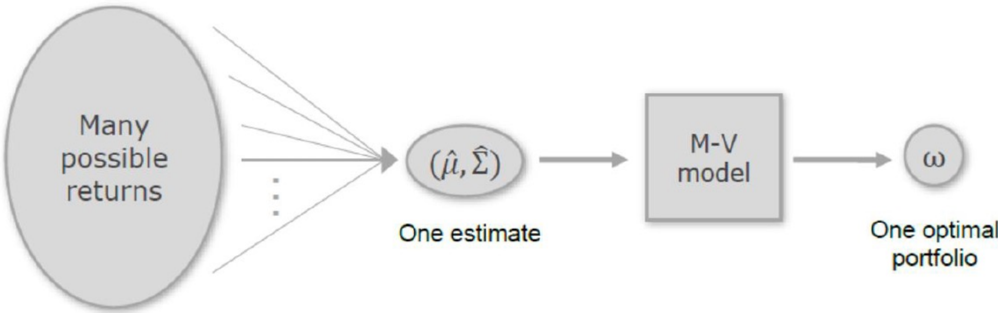
Importance of parameter estimation

- **Accurately estimating the inputs (mean, variance, covariance of asset returns) is challenging**

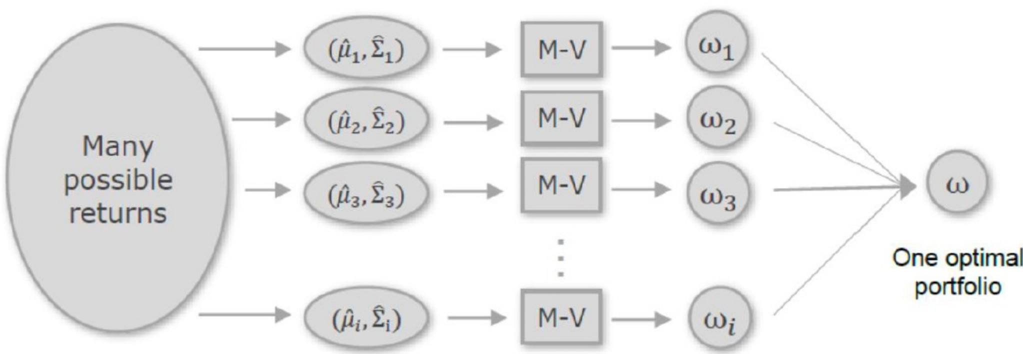
- **There are several approaches to this problem**
  - Constrain weights (upper bound, L_2-norm, …)
  - Improve estimates
  - Portfolio resampling
  - Robust optimization
  - Stochastic programming

Importance of parameter estimation

- **Basic concepts of various robust models**
  - Constraining weights, robust estimators, B-L model

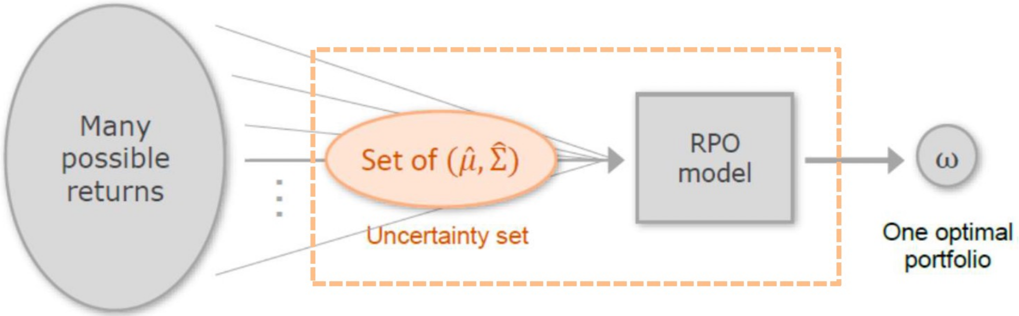


  - Portfolio resampling



Importance of parameter estimation

- **Basic concepts of various robust models**
  - Robust optimization



Power of machine learning

- Traditional quantitative analysis

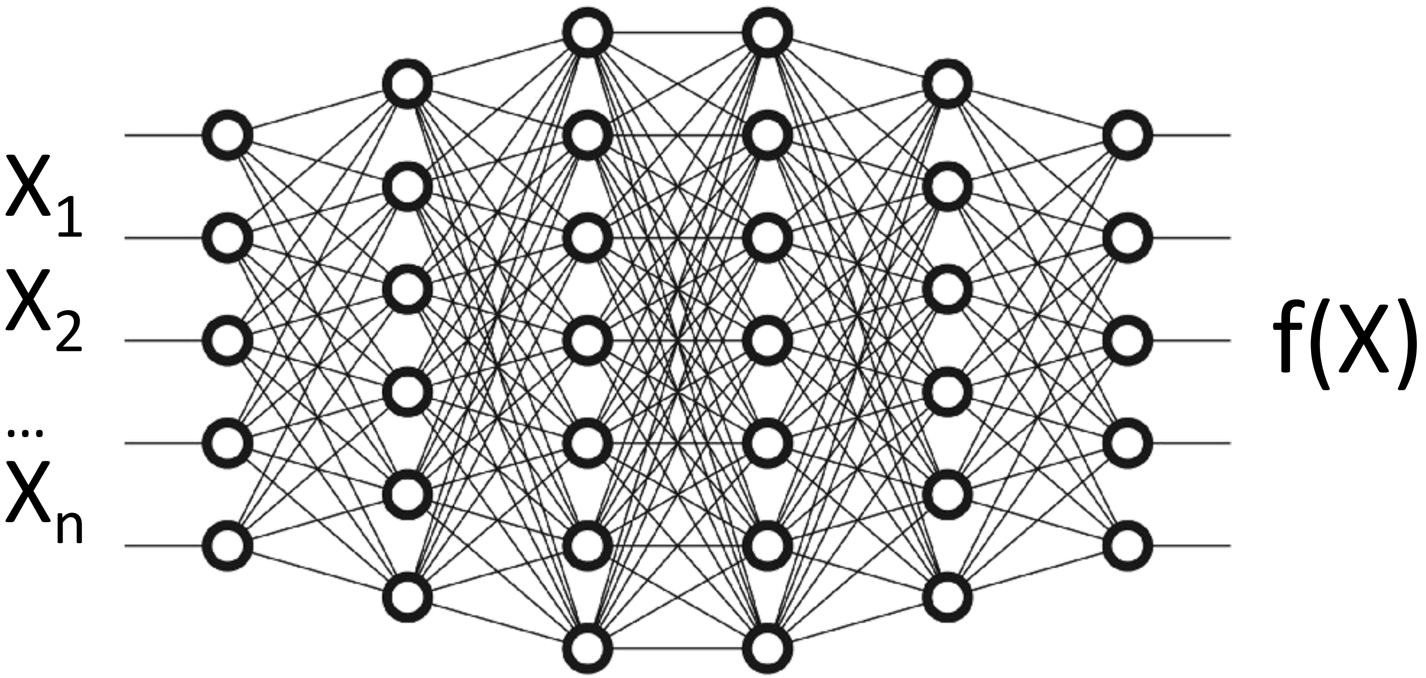
$$f(X) = a+b_1X_1+ b_2X_2+\ldots+ b_nX_n + e$$

Intuitive and explainable

But most relations are nonlinear in real-world

Power of machine learning

- Machine learning



Difficult to interpret

But can learn very complex relationships

Power of machine learning

- Machine learning in finance
  - Empirical asset pricing or factor analysis
    - Litterman and Scheinkman (1991), Kelly, Pruitt, and Su (2019), Lettau and Pelger (2020), Gu, Kelly, and Xiu (2020), Cong et al. (2021), Chen, Pelger, and Zhu (2023), Kelly, Malamud, and Pedersen (2023)
  - Hedging and high-frequency trading
    - Buehler et al. (2019), Buehler, Murray, and Wood (2022)
  - High-frequency trading
    - Sirignano (2019), Ning, Lin, and Jaimungal (2021) and Fang et al. (2021), Cont et al. (2023)
  - Synthetic data generation
    - Polturu et al. (2023) (https://arxiv.org/abs/2401.00081)

  - More applications of machine learning in asset management including above references are well summarized in Lee et al. (2023)

Lee, Yongjae; Thompson, John R.J.; Kim, Jang Ho; Kim, Woo Chang; Fabozzi, Francesco A. (2023) "An Overview of Machine Learning for Asset Management," *Journal of Portfolio Management*, 49(9), 31-63

In this talk

- Two ideas

  – 1. Similarity learning and its application on portfolio optimization

    Hwang, Yoontae; Lee, Junhyeong; Kim, Daham; Noh, Seunghwan; Hong, Joohwan*; Lee, Yongjae* (2023)
    "SimStock: Representation Model for Stock Similarities," *4th ACM International Conference on AI in Finance (ICAIF'23)*, oral presentation (top 27%)

    Hwang, Yoontae; Zohren, Stefan; Lee, Yongjae* (2024)
    "Temporal Representation Learning for Stocks and Its Applications on Investment Management," *working paper*

  – 2. Dynamic robust portfolio optimization
  via Generative Adversarial Networks (GANs)

    Kim, Seyoung; Hong, Joohwan*; Lee, Yongjae* (2023)
    "A GANs-based Approach for Stock Price Anomaly Detection and Investment Risk Management,"
    4th ACM International Conference on AI in Finance (ICAIF'23), oral presentation (top 27%)

    Kim, Jang Ho; Kim, Seyoung; Kim, Woo Chang; Fabozzi, Frank J.; Lee, Yongjae* (2024)
    "Data-Driven Dynamic Robust Portfolio Optimization via Generative Adversarial Networks (GANs)," *working paper*

Section 1

SimStock and Its Application on Portfolio Optimization

Hwang, Yoontae; Lee, Junhyeong; Kim, Daham; Noh, Seunghwan; Hong, Joohwan*; Lee, Yongjae* (2023)
"SimStock: Representation Model for Stock Similarities," *4th ACM International Conference on AI in Finance (ICAIF'23)*, oral presentation (top 27%)

Hwang, Yoontae; Zohren, Stefan; Lee, Yongjae* (2024)
"Temporal Representation Learning for Stocks and Its Applications on Investment Management," *working paper*

Representation learning for stocks

- Traditional categorization of stocks
  - Country
  - Industry sector
  - Style (e.g., size, value)

- They do not work well anymore due to
  - Globalization (e.g., TSMC, Samsung)
  - Digitalization (e.g., Amazon, Tesla)

- We need a more comprehensive approach to define similarity of stocks

SimStock

- SimStock: Representation Learning for Stock Similarities

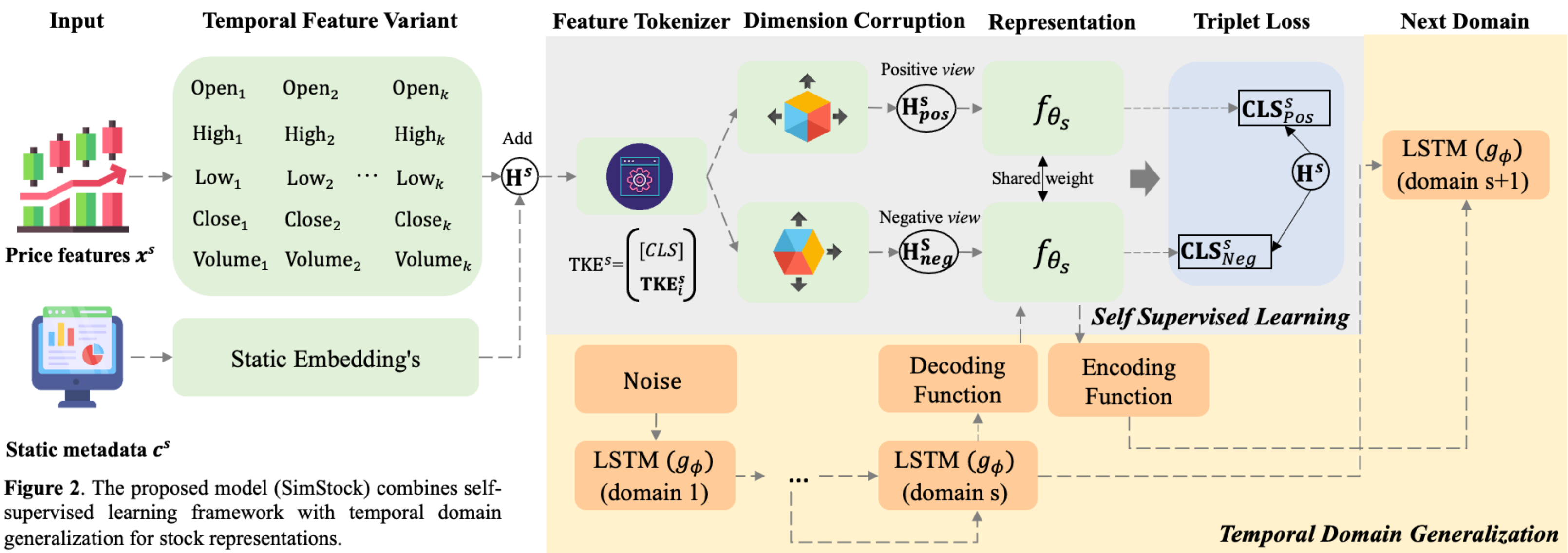
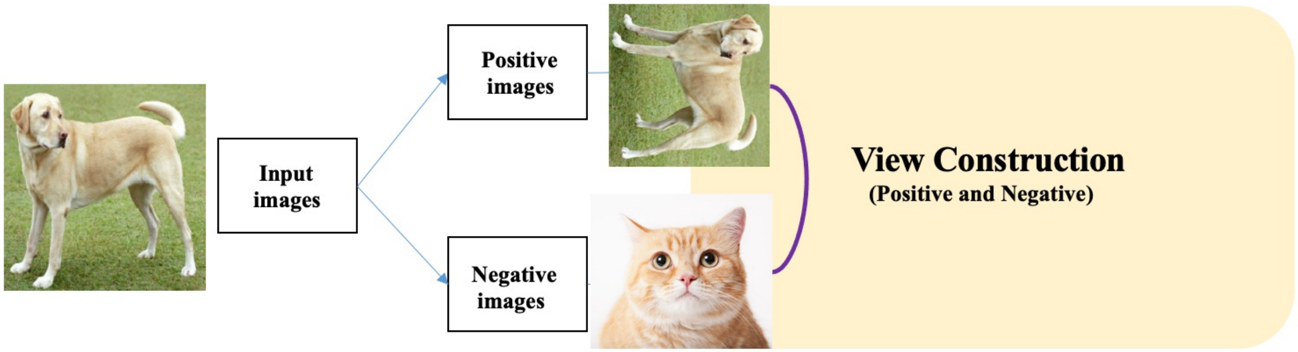


Figure 2. The proposed model (SimStock) combines self-supervised learning framework with temporal domain generalization for stock representations.

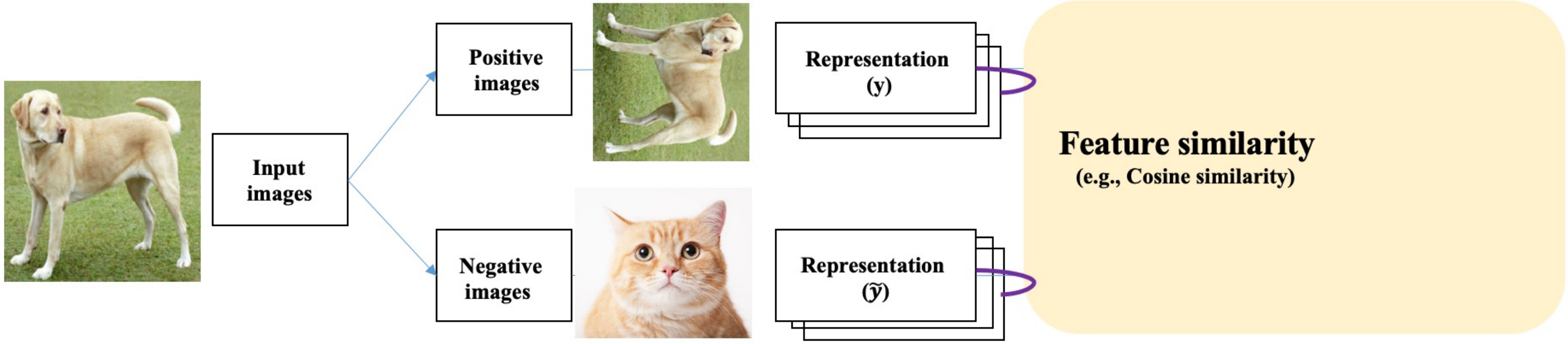
  - Self-supervised learning (SSL)
  - Temporal domain generalization

Hwang, Yoontae; Lee, Junhyeong; Kim, Daham; Noh, Seunghwan; Hong, Joohwan*; Lee, Yongjae** (2023)
"SimStock: Representation Model for Stock Similarities," 4th ACM International Conference on AI in Finance (ICAIF'23), oral presentation (top 27%)

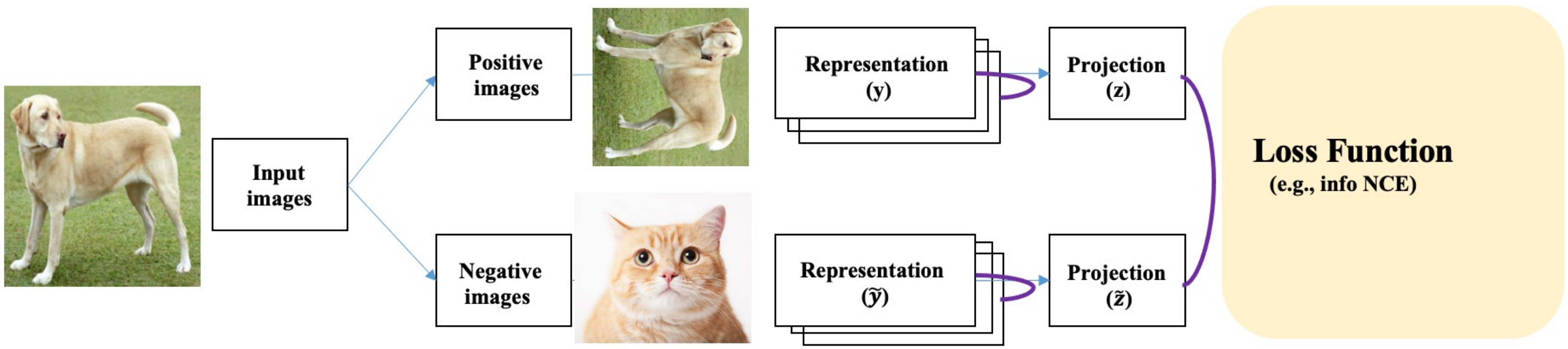
Self-supervised learning



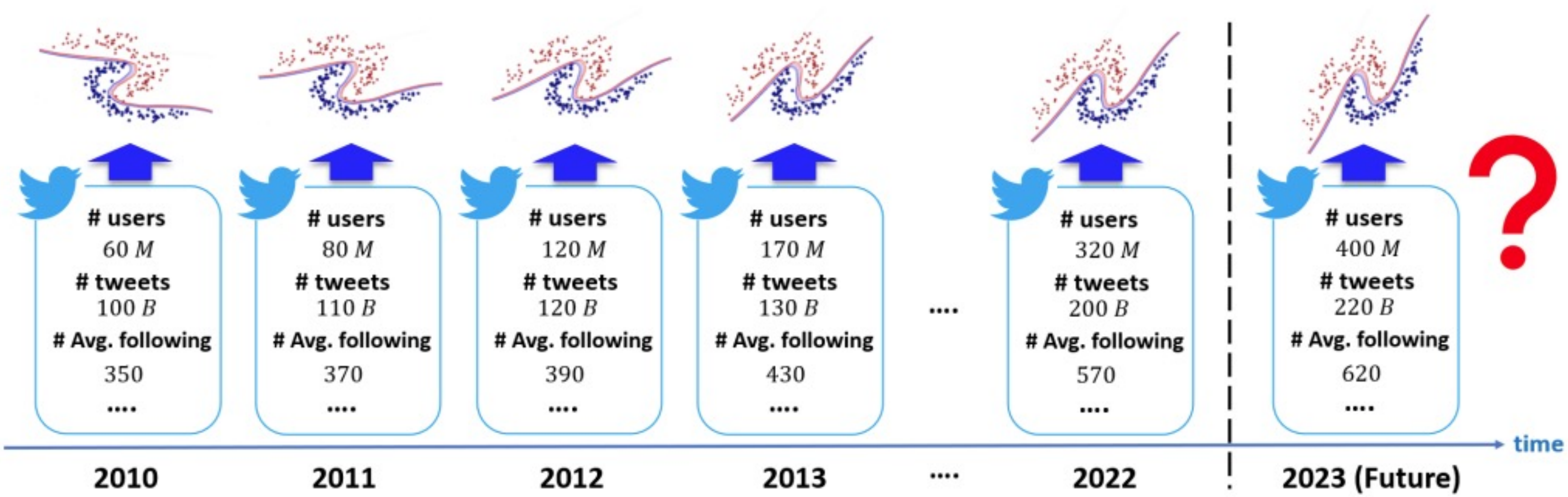
Self-supervised learning



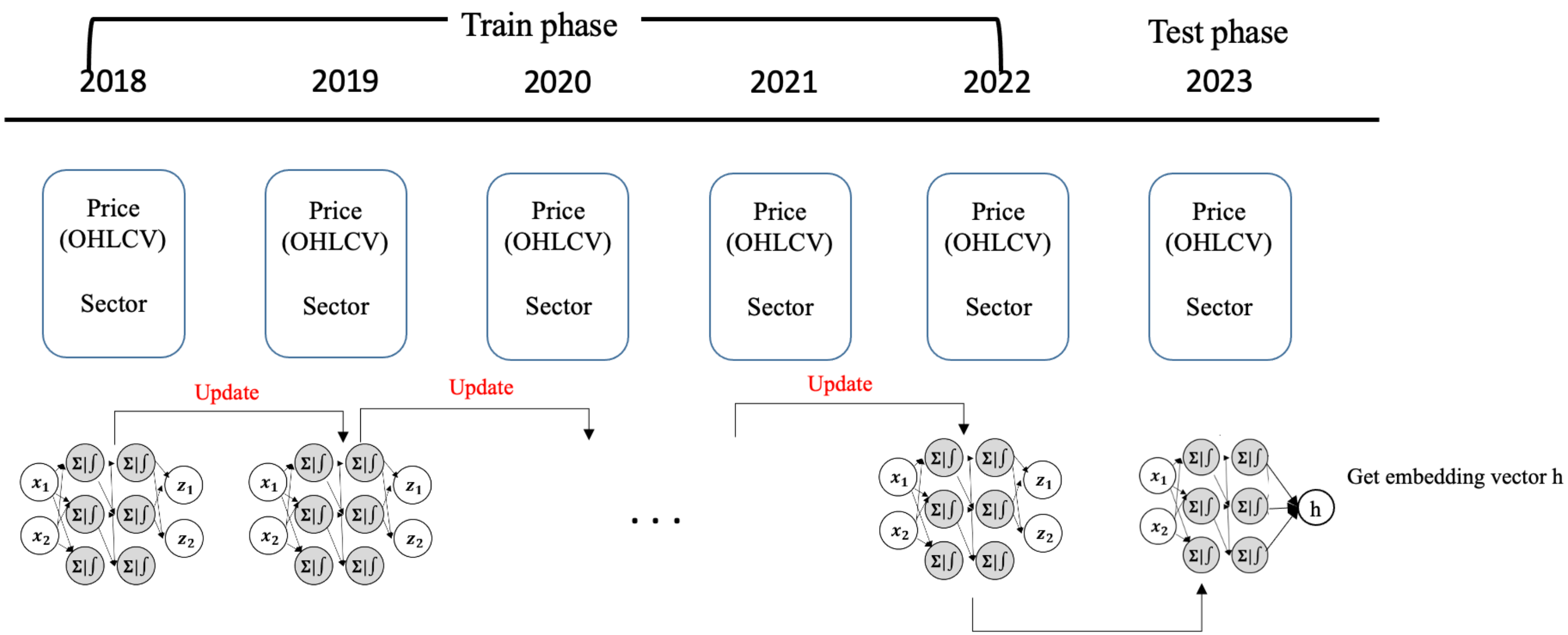
Self-supervised learning



Temporal domain generalization



Temporal domain generalization



SimStock

- SimStock: Representation Learning for Stock Similarities

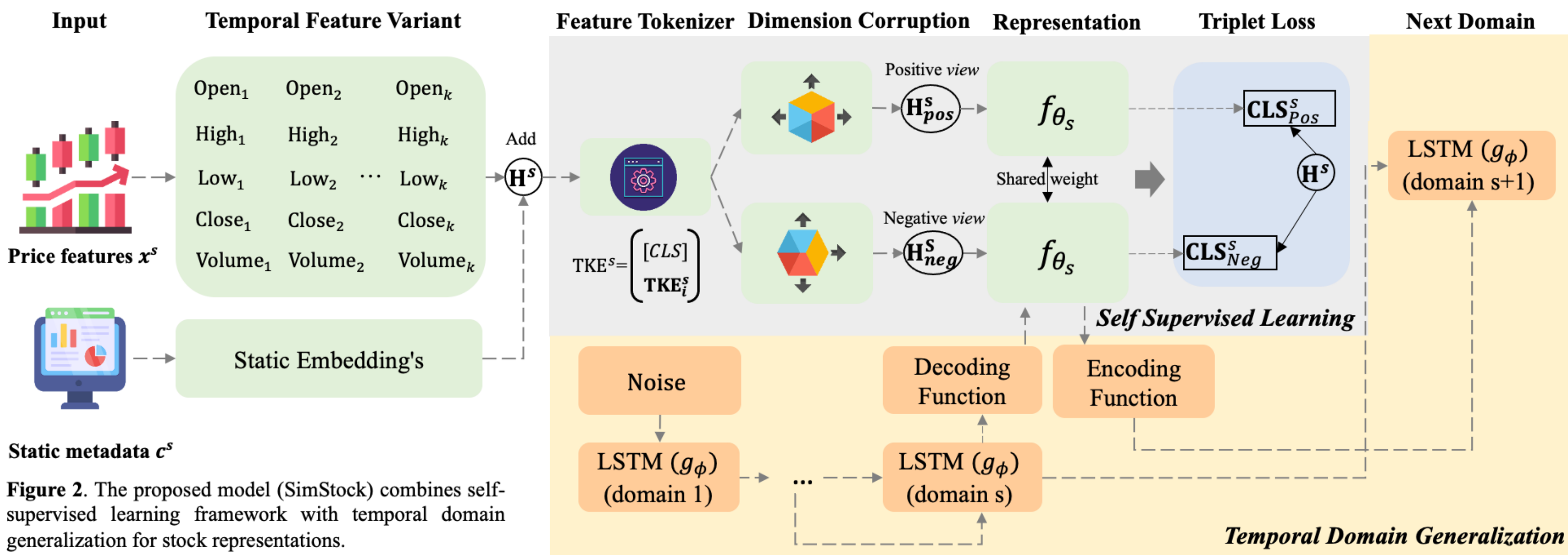


Figure 2. The proposed model (SimStock) combines self-supervised learning framework with temporal domain generalization for stock representations.

  - Self-supervised learning (SSL)
  - Temporal domain generalization

Hwang, Yoontae; Lee, Junhyeong; Kim, Daham; Noh, Seunghwan; Hong, Joohwan*; Lee, Yongjae** (2023)
"SimStock: Representation Model for Stock Similarities," 4th ACM International Conference on AI in Finance (ICAIF'23), oral presentation (top 27%)

Experiment setting

Dataset

NYSE
NASDAQ (US stocks) → 4,231 stocks

SSE(Shanghai Stock exchange) → 1,408 stocks

SZSE(Shenzhen Stock Exchange) → 1,696 stocks

TSE(Tokyo Stock Exchange) → 3,882 stocks

Time period

Training period : Jan 01, 2018 to Dec 31, 2021

Reference period : Jan 01, 2022 to Dec 31, 2022

Test period : Jan 01, 2023 to Dec 31, 2023

Baseline models

Corr1 : past one-year returns correlation

Corr2 : training period returns correlation

Peer : list of similar stocks provided by Google, Yahoo Finance, and Financial Modeling Prep

TS2VEC : Deep learning based state-of-the-art method

Can SimStock find similar stocks?

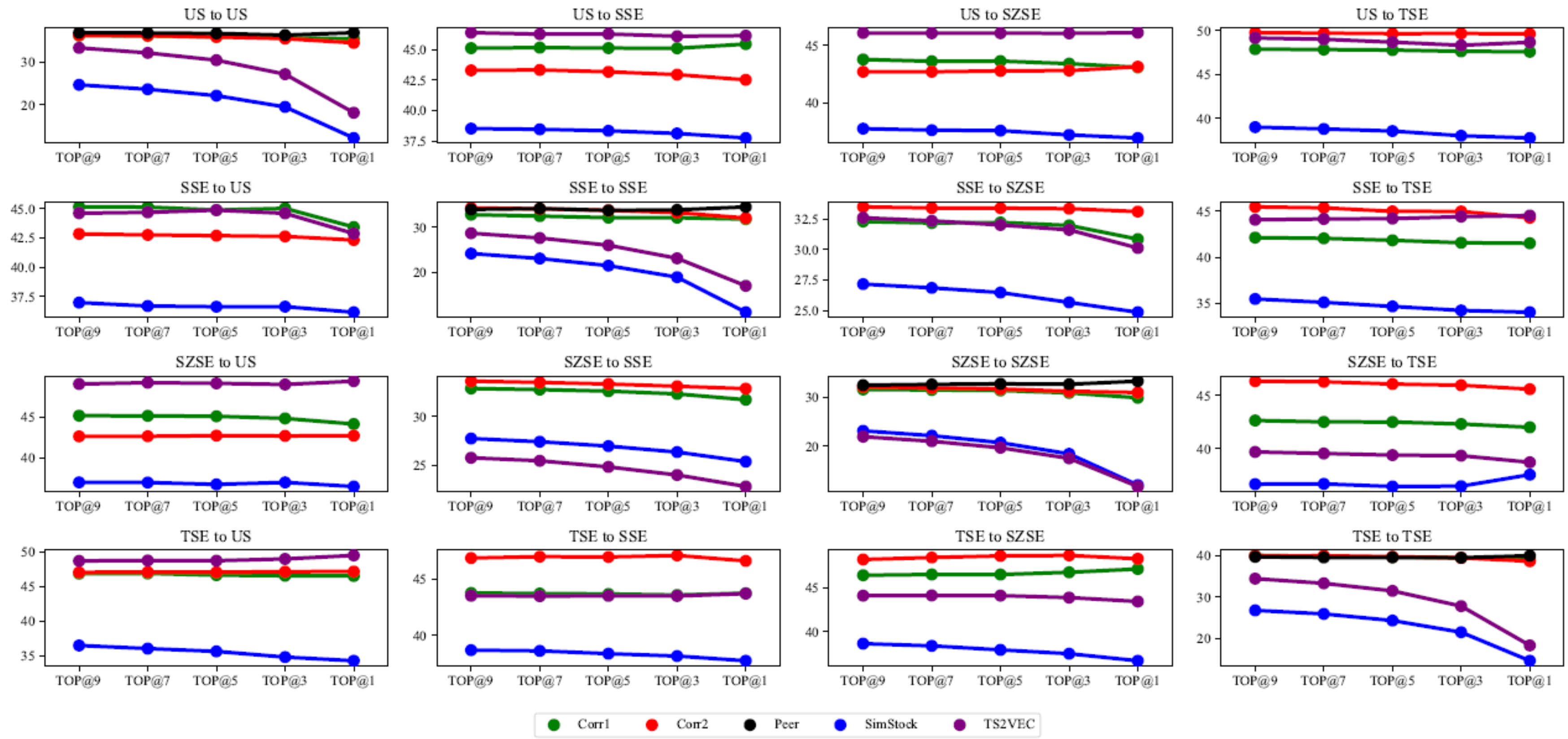


Figure 1. Performance of models in one-to-one and one-to-many scenarios for finding similar stocks.

Application to portfolio optimization

Previous portfolio weights

Portfolio's expected return - Transaction costs

Maximize : $w^T\mu - \psi 1^T |w - w_0|$

Subject to : $w^T \Sigma w \le \sigma^2_{\text{target}}$, Portfolio variance must not exceed predetermined risk target

$w^T 1 = 1$,

$0 \le w_k \le 1$ for all $k = 1,2,..,N$

Application to portfolio optimization

- Benchmark models
  - SimStock Embedding (SS) (ours)
  - Historical Covariance (HC)
  - Shrinkage Method (SM) (Ledoit et al., 2003)
  - Gerber Statistic (GS) (Gerber et al., 2021)
  - TS2VEC (TS) (Yue et al., 2022)

- For SimStock,
  - We use SimStock embeddings to calculate similarity between stocks and replace correlation matrix in portfolio optimization

Application to portfolio optimization

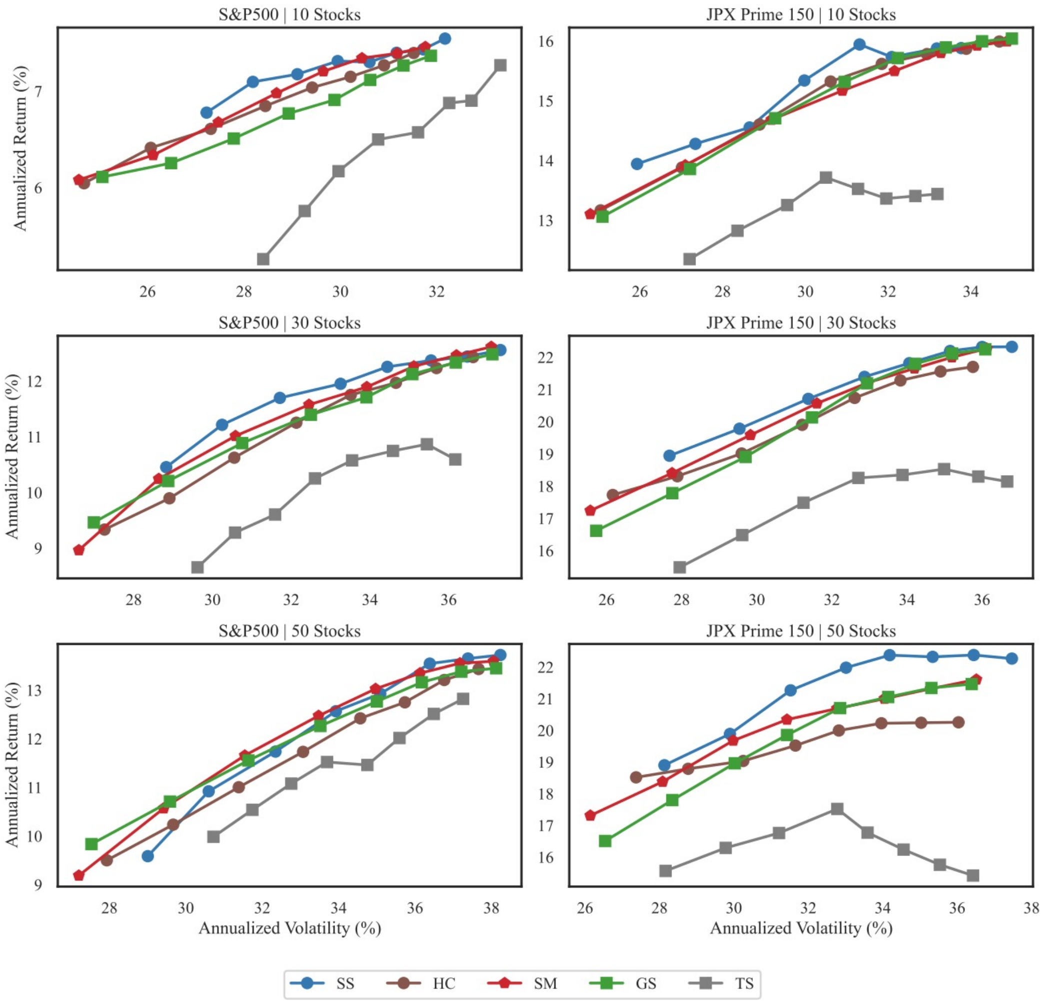


Figure 3 Ex-post efficient frontiers displaying annualized return and volatility of portfolios optimized for different risk targets. The black vertical dotted lines represent the average volatility of the S&P500 and JPX Prime 150, respectively

Application to portfolio optimization

S&P500 — 30 Stocks	Target Volatility (24%)					Target Volatility (27%)					Target Volatility (30%)					Target Volatility (33%)				
Covariance Method	SS	HC	SM	GS	TS	SS	HC	SM	GS	TS	SS	HC	SM	GS	TS	SS	HC	SM	GS	TS
Arithmetic Return (%)	11.31	10.04	9.61	10.14	9.40	12.22	10.75	11.02	11.05	10.51	12.84	11.66	12.03	11.97	10.86	13.33	12.50	12.85	12.69	11.49
Geometric Return (%)	10.39	9.30	8.97	9.43	8.19	11.158	9.86	10.21	10.17	9.43	11.60	10.57	10.967	10.83	9.56	11.88	11.18	11.51	11.32	10.37
Cumulative Return (%)	37.36	33.07	31.60	33.40	26.64	40.54	35.40	36.63	36.49	31.04	42.64	38.44	39.930	39.61	31.53	44.18	41.27	42.59	42.05	34.46
Annualized SD (%)	28.82	27.27	26.61	27.00	29.54	30.24	28.92	28.66	28.90	30.24	31.72	30.58	30.623	30.79	31.37	33.25	32.17	32.49	32.54	32.26
Annualized Skewness	-0.12	-0.15	-0.12	-0.122	-0.17	-0.14	-0.17	-0.16	-0.16	-0.21	-0.18	-0.20	-0.191	-0.20	-0.21	-0.22	-0.22	-0.22	-0.23	-0.24
Annualized Kurtosis	3.17	3.22	3.23	3.161	2.81	3.22	3.28	3.29	3.24	2.86	3.27	3.33	3.343	3.28	2.88	3.27	3.33	3.33	3.28	2.93
Maximum Drawdown (%)	-24.90	-23.96	-23.47	-23.59	-25.59	-25.65	-25.36	-24.57	-25.04	-25.93	-26.69	-26.44	-25.971	-26.57	-26.64	-28.20	-27.63	-27.57	-28.08	-26.47
Monthly 95% VaR (%)	-10.44	-10.22	-9.99	-10.05	-11.11	-10.77	-10.63	-10.53	-10.52	-11.38	-11.19	-10.95	-10.921	-11.01	-11.69	-11.72	-11.40	-11.48	-11.56	-12.2
Sharpe Ratio	0.44	0.40	0.39	0.42	0.31	0.46	0.41	0.43	0.43	0.31	0.46	0.42	0.447	0.43	0.35	0.45	0.43	0.44	0.43	0.36
Annualized Turnover	8.68	8.39	8.49	8.36	7.92	8.69	8.48	8.56	8.49	8.04	8.73	8.54	8.592	8.56	7.99	8.67	8.57	8.51	8.54	7.86

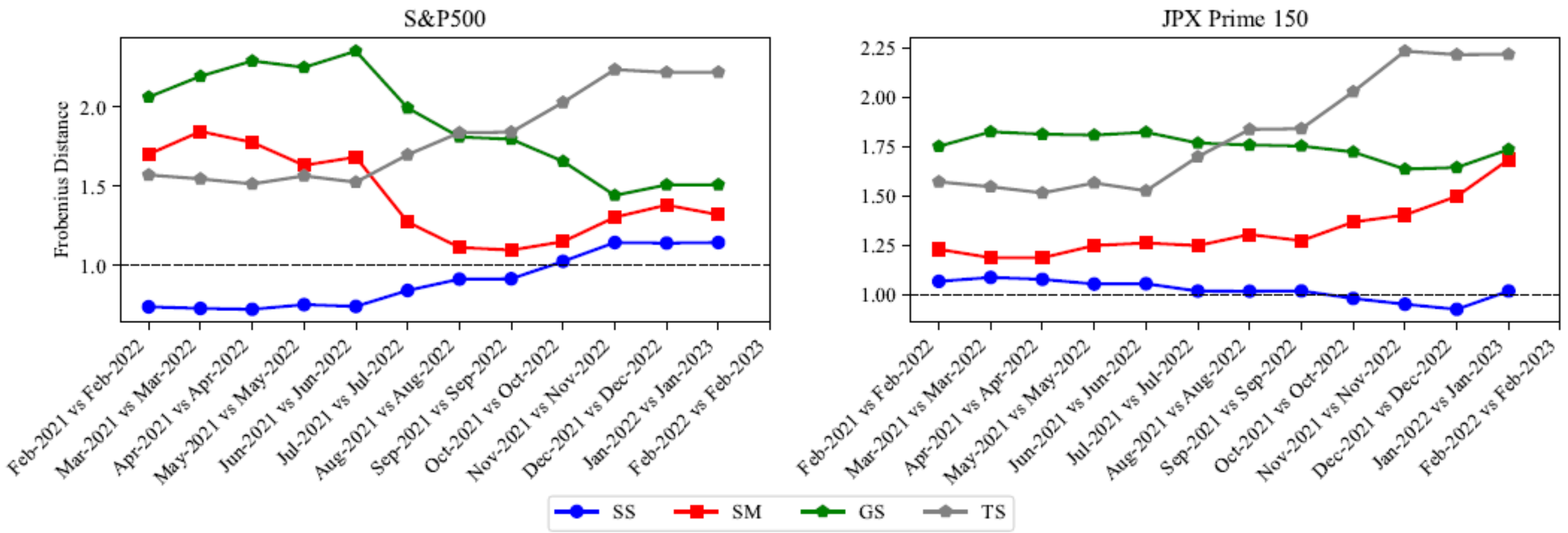
Table 5. This table presents the performance metrics for four portfolio construction methods in the S&P500: Simstock embedding (SS), historical covariance (HC), shrinkage method (SM), Gerber statistic (GS) and TS2VEC embedding (TS). The portfolios were optimized for four different risk target levels: 24%, 27%, 30%, and 33%. The performance was evaluated over the full testing period from January 2022 to February 2024. The 3-month U.S. Treasury Bill rate was used as the risk-free rate for performance calculations. Transaction costs were modeled as 10 basis points of the traded volume for each rebalancing event.

JPX Prime 150 — 30 Stocks	Target Volatility (24%)					Target Volatility (27%)					Target Volatility (30%)					Target Volatility (33%)				
Covariance Method	SS	HC	SM	GS	TS	SS	HC	SM	GS	TS	SS	HC	SM	GS	TS	SS	HC	SM	GS	TS
Arithmetic Return (%)	20.12	19.02	18.61	17.80	14.60	20.85	19.78	20.02	19.20	15.67	21.70	20.65	21.40	20.51	16.83	22.47	21.70	22.50	21.96	18.45
Geometric Return (%)	18.63	17.72	17.24	16.62	12.96	19.24	18.31	18.40	17.78	14.25	19.96	19.01	19.58	18.90	15.19	20.62	19.90	20.57	20.14	16.79
Cumulative Return (%)	71.15	66.53	64.60	61.55	44.16	74.32	69.56	70.25	67.23	49.12	77.95	73.23	76.27	72.90	52.85	81.67	77.87	81.37	79.30	59.32
Annualized SD (%)	26.83	26.16	25.56	25.71	27.19	28.51	27.88	27.74	27.74	28.39	29.99	29.59	29.83	29.69	29.62	31.33	31.20	31.59	31.46	30.66
Annualized Skewness	0.28	0.16	0.17	0.13	0.11	0.32	0.17	0.21	0.18	0.15	0.34	0.19	0.23	0.21	0.16	0.32	0.19	0.23	0.20	0.13
Annualized Kurtosis	3.37	2.94	2.99	2.92	2.69	3.43	3.02	3.08	3.01	2.67	3.49	3.06	3.17	3.09	2.73	3.51	3.12	3.22	3.16	2.72
Maximum Drawdown (%)	-19.17	-19.63	-19.52	-19.44	-21.87	-20.37	-20.89	-20.85	-20.62	-22.89	-21.16	-22.09	-22.15	-21.96	-22.57	-22.19	-23.28	-23.29	-23.14	-24.30
Monthly 95% VaR (%)	-8.52	-8.89	-8.49	-8.82	-9.86	-9.00	-9.46	-9.20	-9.36	-9.92	-9.41	-9.95	-9.79	-9.90	-10.29	-9.79	-10.46	-10.31	-10.40	-10.75
Sharpe Ratio	0.96	0.92	0.91	0.86	0.63	0.93	0.89	0.90	0.86	0.65	0.92	0.87	0.90	0.87	0.67	0.92	0.87	0.90	0.88	0.71
Annualized Turnover	8.81	8.38	8.50	8.54	8.21	8.83	8.52	8.57	8.62	8.15	8.81	8.59	8.56	8.59	8.29	8.85	8.58	8.54	8.54	8.19

Table 6. This table presents the performance metrics for four portfolio construction methods in the JPX Prime 150: Simstock embedding(SS), historical covariance (HC), shrinkage method (SM), Gerber statistic (GS) and TS2VEC embedding (TS). The portfolios were optimized for four different risk target levels: 24%, 27%, 30%, and 33%. The performance was evaluated over the full testing period from January 2022 to February 2024. The 3-month U.S. Treasury Bill rate was used as the risk-free rate for performance calculations. Transaction costs were modeled as 10 basis points of the traded volume for each rebalancing event.

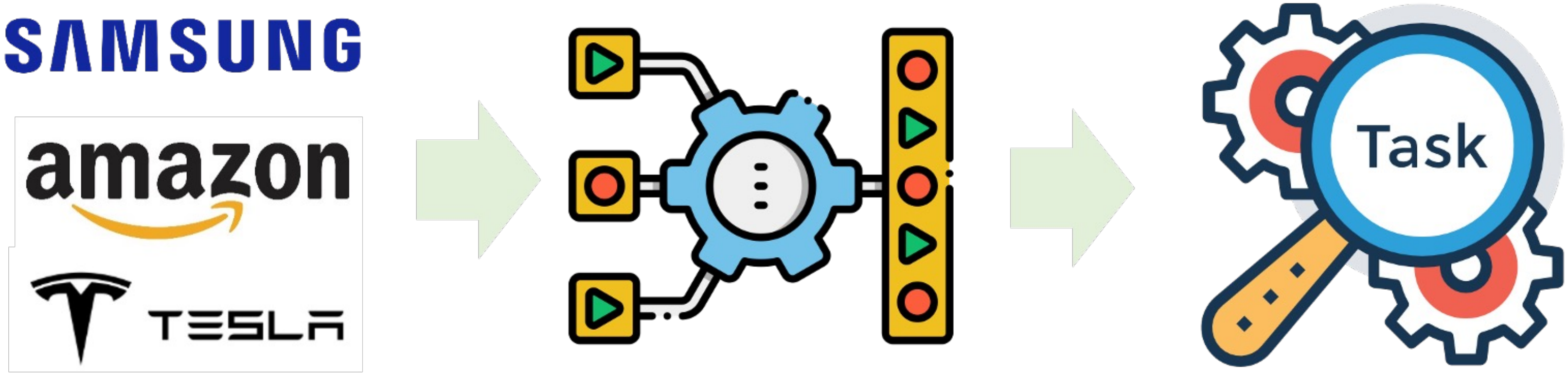
Application to portfolio optimization

$$\frac{\left\|\mathrm{Corr}_X - \mathrm{Corr}^{\mathrm{future}}\right\|_F}{\left\|\mathrm{Corr}_X - \mathrm{Corr}^{\mathrm{past}}\right\|_F}$$



Summary

- SimStock can be useful in modeling stock similarities
- They can be applied to various tasks in investment management



Section 2

Dynamic Robust Portfolio Optimization via Generative Adversarial Networks (GANs)

Kim, Seyoung; Hong, Joohwan*; Lee, Yongjae* (2023)
"A GANs-based Approach for Stock Price Anomaly Detection and Investment Risk Management,"
4th ACM International Conference on AI in Finance (ICAIF'23), oral presentation (top 27%)

Kim, Jang Ho; Kim, Seyoung; Kim, Woo Chang; Fabozzi, Frank J.; Lee, Yongjae* (2024)
"Data-Driven Dynamic Robust Portfolio Optimization via Generative Adversarial Networks (GANs)," *working paper*

MV Optimization

- Modern Portfolio Theory (MPT)
  - Originated from the pioneering work of Markowitz (1952)

$$\min_{w} \ w^T \Sigma w$$
$$\text{s.t.} \quad \mathbf{1}^{\mathrm{T}} w = 1$$
$$w^T \mu \geq \mu_0$$

$w \in \mathbb{R}^n$ = portfolio weight vector,
$\mu \in \mathbb{R}^n$ = mean return vector of n securities,
$\Sigma \in \mathbb{R}^{n \times n}$ = covariance matrix of n securities.

  - Mathematically established the concept of diversification
    - Trade-off relationship between risk and return
  - It has been widely used for asset allocation in practice (Kim et al., 2020)

Limitation of MV Optimization

- Michaud (1989):
  "MV optimization is ***estimation-error maximizers***"

  – MV optimization gives significant weights to securities with large estimated returns, negative correlations and small variances, which most likely have large estimation errors

- **MV optimization is too sensitive** to errors in input parameters

Problems of Error Sensitiveness

- **Unstable optimal solution**
  - Small change in inputs can cause large changes in the solutions
  - Klein and Bawa (1976)
    - Assume stock return has normal distribution with parameter θ
    - MV optimal solution under estimated θ differ from unconditional MV optimal solution
  - Best and Grauer (1991)
    - Errors in mean return significantly change the composition of optimal portfolio

Problems of Error Sensitiveness

- **Poor performance**
  - Jobson and Korkie (1980)
    - Equal weight portfolio might outperform MV optimal portfolios due to estimation errors
  - Broadie (1993)
    - The actual performances of optimal portfolios with estimated parameters are below those of optimal portfolios with true parameters

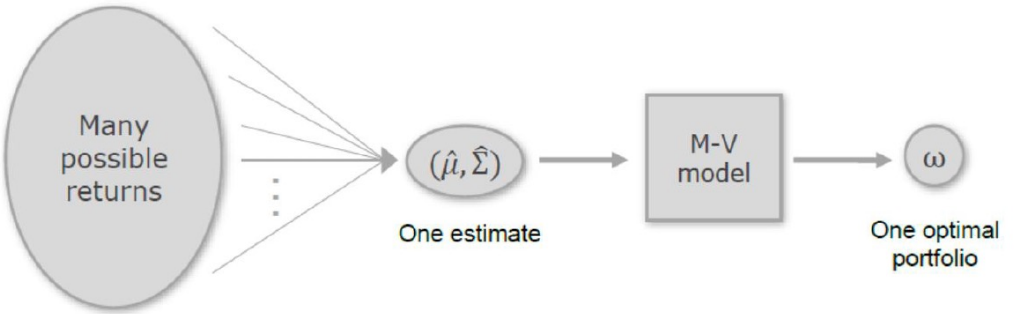
Importance of parameter estimation

- **Accurately estimating the inputs (mean, variance, covariance of asset returns) is challenging**

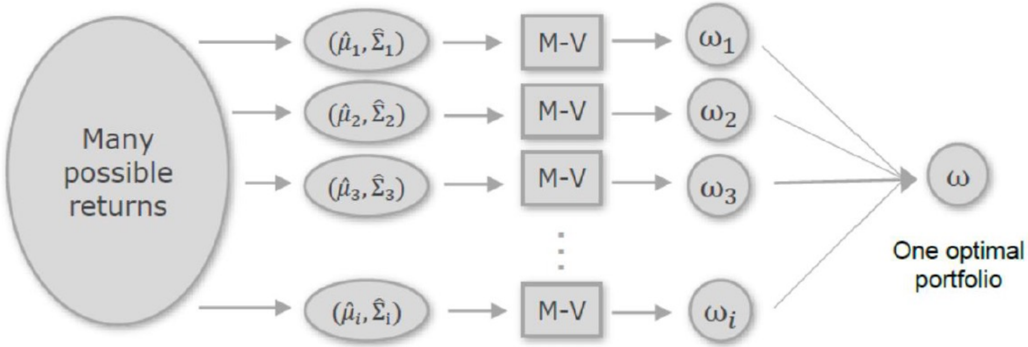
- **There are several approaches to this problem**
  - Constrain weights (upper bound, L_2-norm, …)
  - Improve estimates
  - Portfolio resampling
  - **Robust optimization**
  - Stochastic programming

Importance of parameter estimation

- **Basic concepts of various robust models**
  - Constraining weights, robust estimators, B-L model

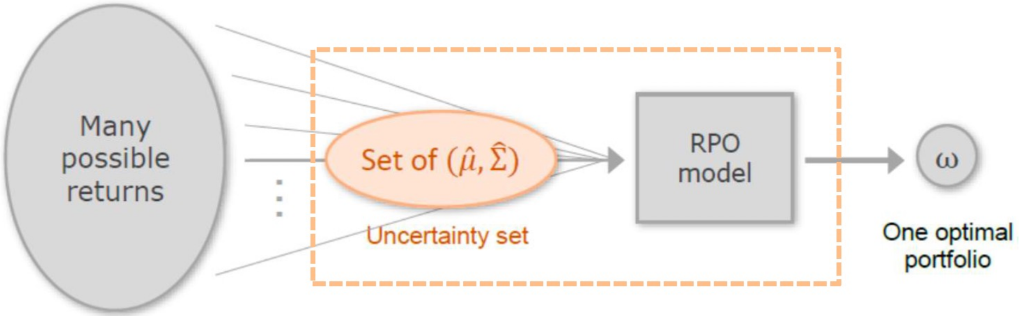


  - Portfolio resampling



Importance of parameter estimation

- **Basic concepts of various robust models**
  - Robust optimization



Robust optimization

- Robust **optimization**

  – Constructing results that are **insensitive to small deviations** from the model assumptions (or input parameters)

  – First established in 1950s. Since then, it has been applied in statistics, operations research, electrical engineering, control theory, finance, logistics, etc.

Markowitz model

- Markowitz mean-variance portfolio optimization

$$\min_{w \in \Omega} \; w^T \boxed{\Sigma} w - \lambda w^T \boxed{\mu}$$

$w \in \mathbb{R}^n$: portfolio weight
$\Sigma \in \mathbb{R}^{n \times n}$: covariance matrix
$\mu \in \mathbb{R}^n$: expected return
$\lambda \in \mathbb{R}$: risk-preference
$\Omega \subseteq \mathbb{R}^n$: set of feasible portfolios

– Simplest feasible set would be
  - $\Omega = \{w \in \mathbb{R}^n \mid w^T \mathbf{1} = 1\}$
  - One may add more constraints

– In reality, it is difficult to accurately measure mean, variance, and covariance of asset returns
  - **$\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are uncertain parameters**

Robust portfolio optimization

- Robust portfolio optimization

worst-case

$$\min_{w\in\Omega} \boxed{\max_{(\mu,\Sigma)\in\mathcal{U}} w^T\Sigma w - \lambda w^T\mu}$$

$w \in \mathbb{R}^n$: portfolio weight
$\Sigma \in \mathbb{R}^{n\times n}$: covariance matrix
$\mu \in \mathbb{R}^n$: expected return
$\lambda \in \mathbb{R}$: risk-preference
$\Omega \subseteq \mathbb{R}^n$: set of feasible portfolios
$\mathcal{U} \subseteq (\mathbb{R}^n, \mathbb{R}^{n\times n})$: possible values for μ and Σ

– Find the best solution among the solutions that satisfy the constraints for all realizations of the **uncertain** components

– That is, find a solution that has the best solution under its **worst case**

Uncertainty sets

- Assume that an uncertain parameter has its value within a set
  - And it is called an **uncertainty set**
  - Below are some examples of uncertainty sets of μ
    - Scenario uncertainty: $\{\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(s)}\}$
    - Box (interval) uncertainty: $\{\mu \mid |\mu_{\mathrm{i}} - \hat{\mu}_i| \leq \delta_i, i = 1, \dots, n\}$
    - Ball uncertainty: $\{\mu \mid \|\mu - \hat{\mu}\|_2 \leq \delta\}$
    - Ellipsoidal uncertainty: $\{\mu \mid \left\|\Sigma_\mu^{-\frac{1}{2}}(\mu - \hat{\mu})\right\|_2 \leq \delta\}$
    - Budget uncertainty: $\{\mu \mid \|\mu - \hat{\mu}\|_\infty \leq \delta, \|\mu - \hat{\mu}\|_1 \leq \gamma\}$

Robust portfolios under box uncertainty

- Uncertainty set $\mathcal{U}_\delta(\hat{\mu}) = \{\mu \mid |\mu_i - \hat{\mu}_i| \le \delta_i, i = 1, \dots, n\}$

$$\min_{w \in \Omega} \max_{\mu \in \mathcal{U}_\delta(\hat{\mu})} w^T \Sigma w - \lambda w^T \mu$$

$$\Downarrow$$

$$\min_{w \in \Omega} w^T \Sigma w - \lambda(\hat{\mu}^T w - \delta^T |w|)$$

$w \in \mathbb{R}^n$: portfolio weight
$\Sigma \in \mathbb{R}^{n \times n}$: covariance matrix
$\mu \in \mathbb{R}^n$: expected return
$\lambda \in \mathbb{R}$: risk-preference
$\Omega \subseteq \mathbb{R}^n$: set of feasible portfolios

  – Normality of asset returns can be assumed (or by CLT) for computing the value of δ

  – Separate bound can be set for each asset

Robust portfolios under box uncertainty

- Uncertainty set $\mathcal{U}_\delta(\hat{\mu}) = \{\mu \mid |\mu_i - \hat{\mu}_i| \le \delta_i, i = 1, \dots, n\}$

$$\min_{w \in \Omega} \max_{\mu \in \mathcal{U}_\delta(\hat{\mu})} w^T \Sigma w - \lambda w^T \mu$$

$$\Downarrow$$

$$\min_{w \in \Omega} w^T \Sigma w - \lambda(\hat{\mu}^T w - \delta^T |w|)$$

$$\Downarrow$$

$$\min_{w \in \Omega} w^T \Sigma w - \lambda \hat{\mu}^T w + \lambda \delta^T (w_+ + w_-)$$

$$s.t. \;\; w = w_+ - w_-$$

$$w_+ \ge 0, \;\; w_- \ge 0$$

$w \in \mathbb{R}^n$: portfolio weight
$\Sigma \in \mathbb{R}^{n \times n}$: covariance matrix
$\mu \in \mathbb{R}^n$: expected return
$\lambda \in \mathbb{R}$: risk-preference
$\Omega \subseteq \mathbb{R}^n$: set of feasible portfolios

– Note that it is a quadratic programming problem

Robust portfolios under box uncertainty

- Uncertainty set $\mathcal{U}_\delta(\hat{\mu}) = \{\mu \mid |\mu_i - \hat{\mu}_i| \le \delta_i, i = 1, \dots, n\}$

$$\min_{w\in\Omega} \max_{\mu\in\mathcal{U}_\delta(\hat{\mu})} w^T\Sigma w - \lambda w^T \mu$$

$w \in \mathbb{R}^n$: portfolio weight
$\Sigma \in \mathbb{R}^{n\times n}$: covariance matrix
$\mu \in \mathbb{R}^n$: expected return
$\lambda \in \mathbb{R}$: risk-preference
$\Omega \subseteq \mathbb{R}^n$: set of feasible portfolios

$$\min_{w\in\Omega} w^T\Sigma w - \lambda(\hat{\mu}^T w - \delta^T |w|)$$

$$\min_{w\in\Omega} w^T\Sigma w - \lambda\hat{\mu}^T w + \lambda\delta^T(w_+ + w_-)$$

$$s.t.\ \ w = w_+ - w_-$$

$$w_+ \ge 0,\ \ w_- \ge 0$$

Then, how can we set the values of δ_i?

Also, it should be *time-varying* considering the huge volatility in financial markets.

– Note that it is a quadratic programming problem

Research objective

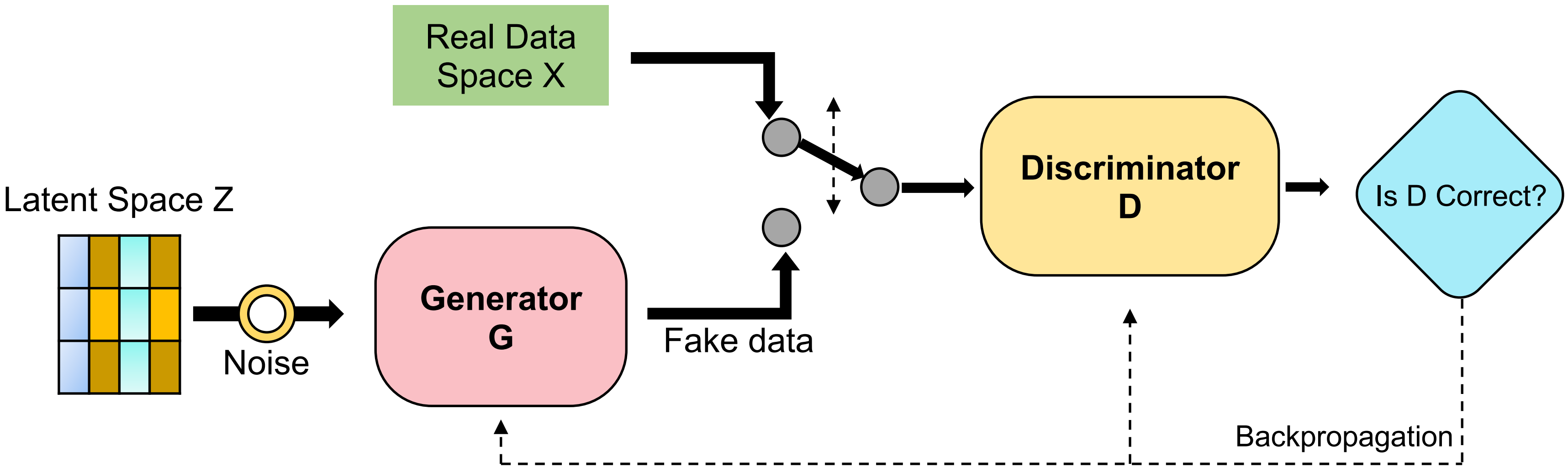
- Develop a robust portfolio optimization model
  - That can adjust the size of uncertainty set dynamically
  - Using a data-driven approach

Research objective

- Develop a robust portfolio optimization model
  - That can adjust the size of uncertainty set dynamically
  - Using a **data-driven approach**

*In this regard, we use **Generative Adversarial Networks (GANs)***

Generative Adversarial Networks

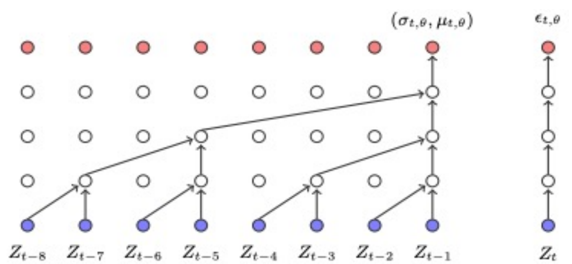


- Generative Adversarial Network (GAN) is an advanced generative model that creates realistic synthetic data by competing two neural networks against each other in an adversarial setting.
- Generator, a neural network, creates synthetic data by mapping from a latent space to the training data distribution. It aims to generate data that resembles real data and deceive the discriminator.
- Discriminator, another neural network, distinguishes between synthetic data from the generator and actual data, determining the origin of input data.

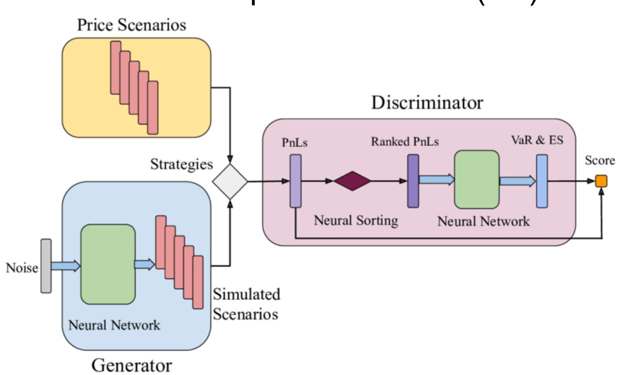
GANs in Finance

Generating Financial Data

- **Quant GANs (Wiese et al, 2020)**
  - Generate financial time series dependence characteristics via TCN network-based GANs.

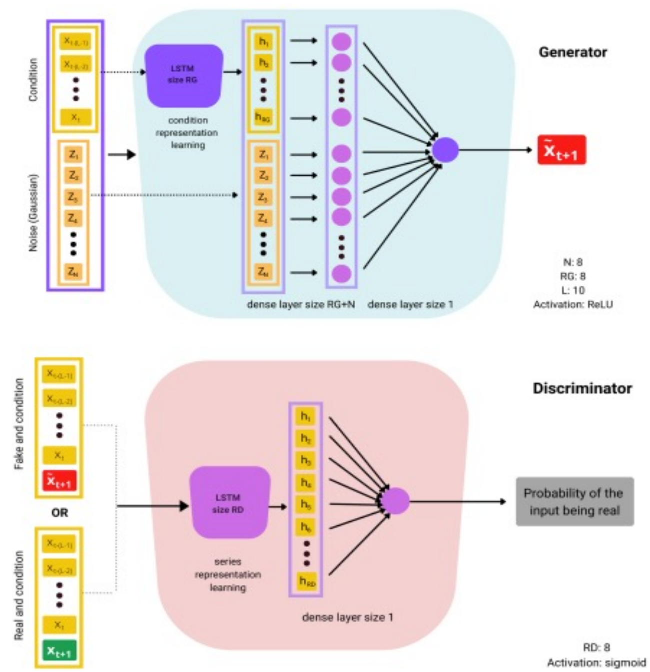


- **Tail-GAN (Cont et al, 2022)**
  - Generate market scenarios reflecting risk measures such as VaR and expected shortfall (ES).



Predicting Stock Price

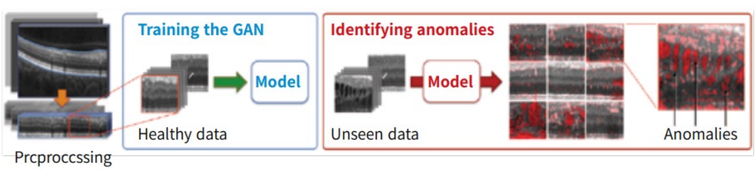
- **Fin-GAN (Vuletić et al, 2023)**
  - Modify the Generator's loss function, transforming GANs into a supervised learning framework for classifying the direction of stock price movements.

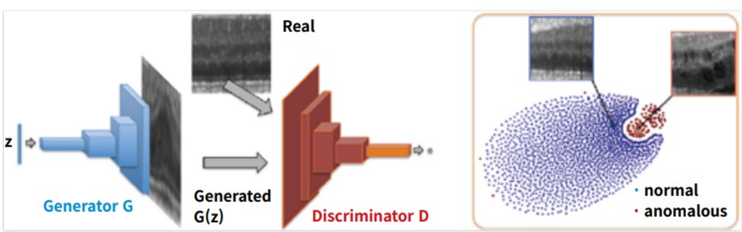


GAN-based Anomaly Detection

Image

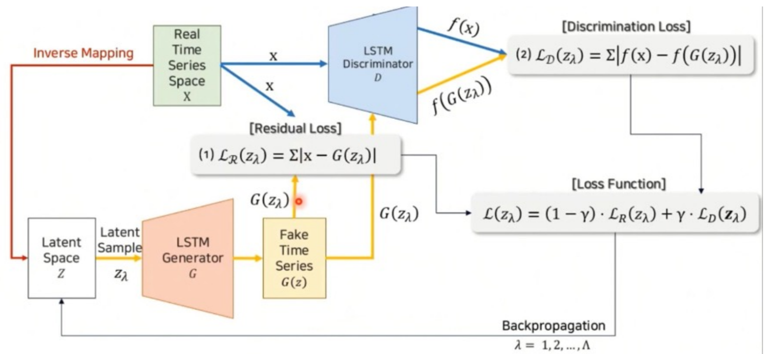
- **AnoGAN (Schlegl et al., 2017)**
  - Difficulty in learning high-dimensional image data
  - Low diversity of generated images



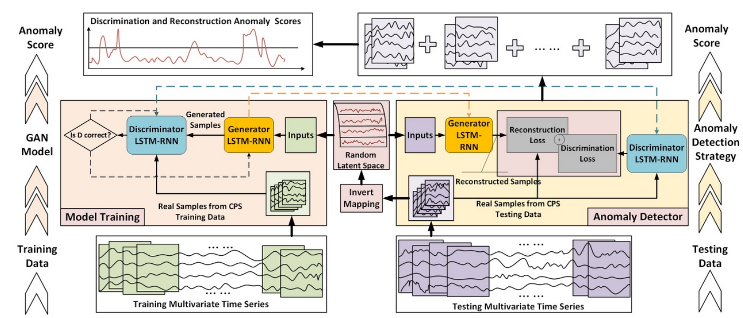


Time-series

- **TAnoGAN (Bashar & Nayak, 2020)**
  - As the input is time-series data, it is used by applying a time window to the sequence



- **MAD-GAN (Li et al., 2020)**
  - Multivariate time-series data



Data

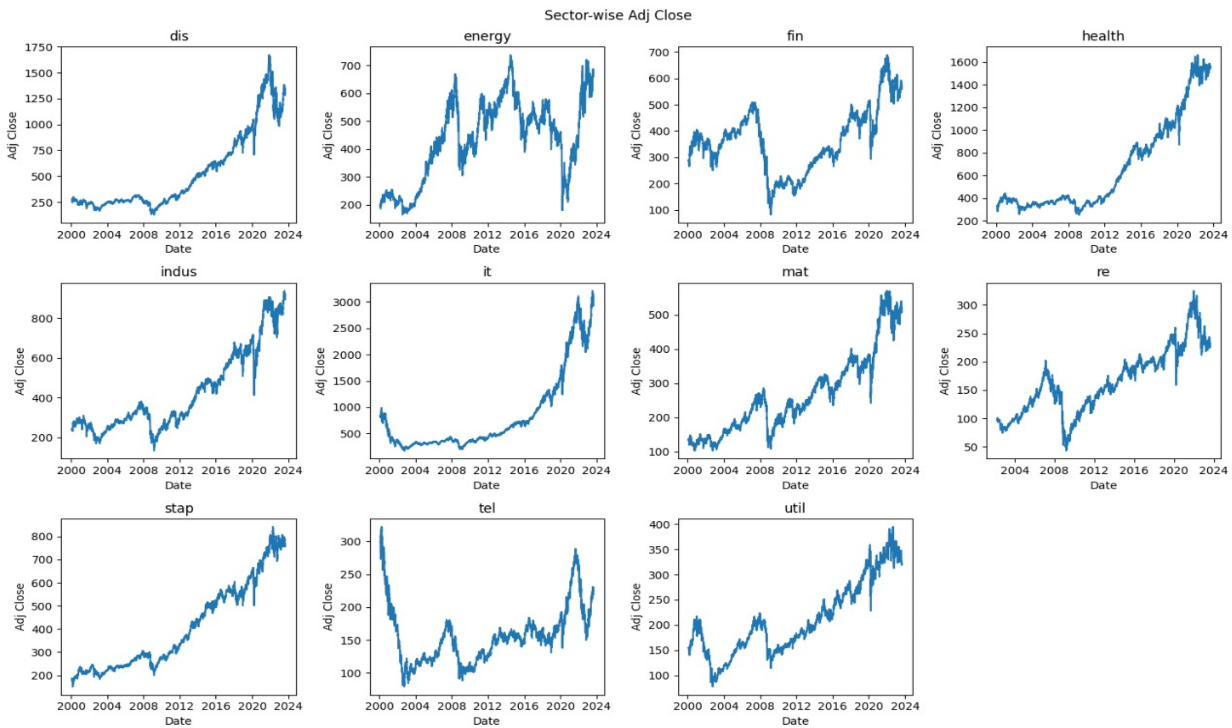
Data description

- Adjusted close price of **S&P500 index** from yahoo finance
- Training period : 2000/01/01 ~ 2019/12/31 (excluding 2007~2009)
- Testing period : 2021/07/01 ~ 2023/08/30 (including global downturn in 2022)
- Validation period : 2007/01/01 ~ 2009/12/31 (including global crisis)
- Anomaly signal lookback periods searching: 2020/01/01 ~2021/06/30 (Including COVID-19)

Data

Data description

- Adjusted close prices of **11 S&P500 industry sector indices** from yahoo finance
- Training period : 2000/01/01 ~ 2019/12/31 (excluding 2007~2009)
- Testing period : 2021/07/01 ~ 2023/08/30 (including global downturn in 2022)
- Validation period : 2007/01/01 ~ 2009/12/31 (including global crisis)
- Anomaly signal lookback periods searching: 2020/01/01 ~2021/06/30 (Including COVID-19)

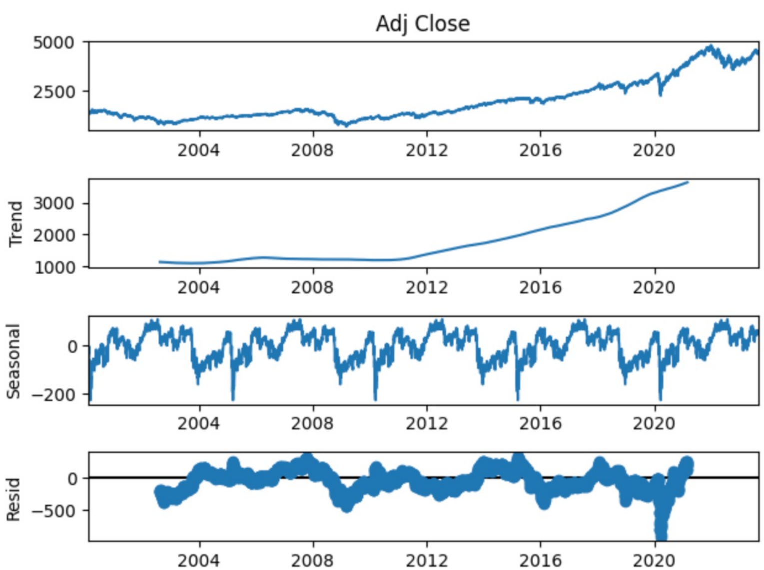


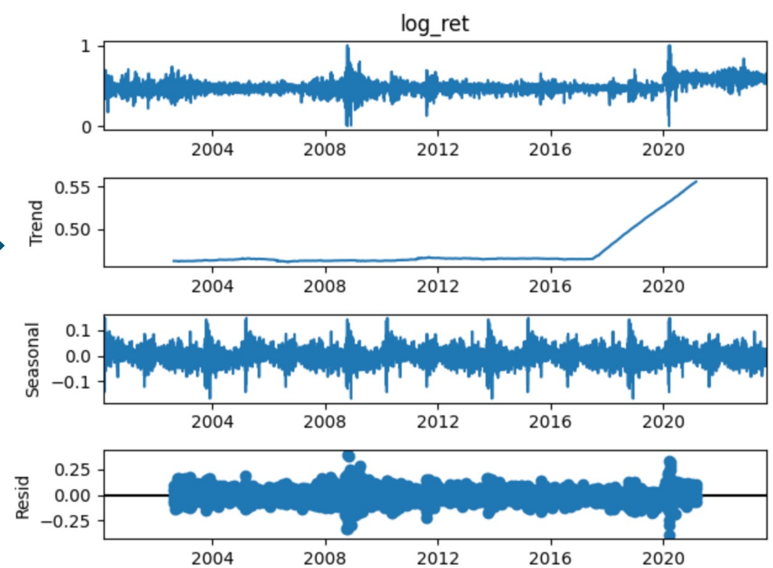
(https://felab.unist.ac.kr)

Data Pre-processing

Log-return

- When working with financial-time series data, such as stock prices, we transform it into log-returns.
- First, it helps to view the data from a stationary perspective.
- Second, it is easier to compute and work with than raw price data.
- Finally, it can help correct for non-linearity in the data, making it a useful tool in financial analysis.



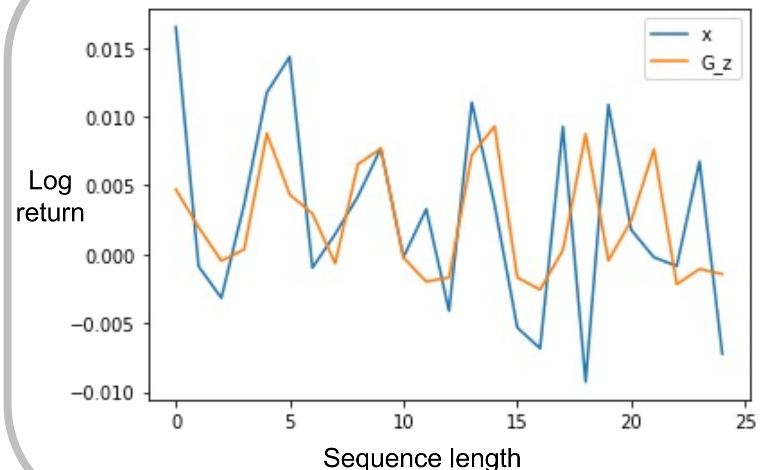


Paths for the generated sequence data

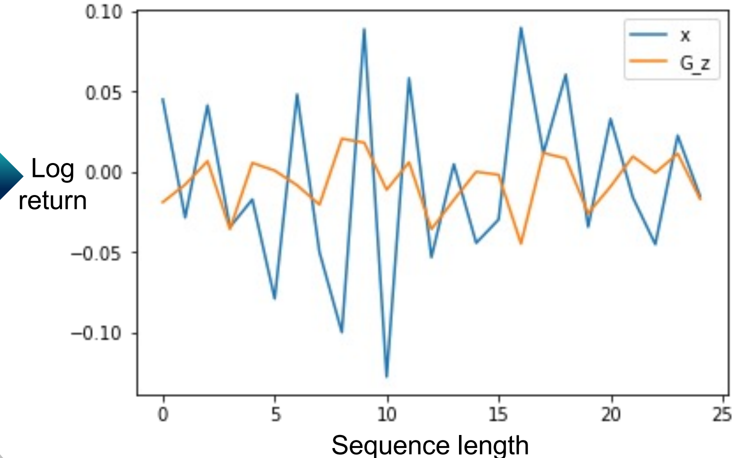
Generated paths

- The latent sample $z_{i,j}$ is updated via 50 backpropagation iterations to make G($z_{i,j}$) as similar as possible to real data x.
- During the sampling of $z_{i,j}$, the process involves resampling 3 times and selecting one with the lowest anomaly loss.
- The generated paths showed similar patterns to the actual data movement during the test period.
- However, paths showed a very different patterns during periods of extreme market instability such as the COVID shock.

Before the COVID-19 pandemic

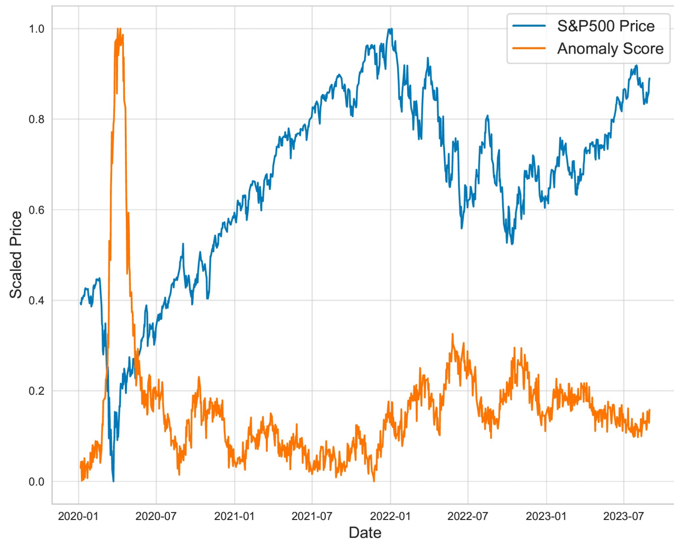


During the COVID-19 pandemic

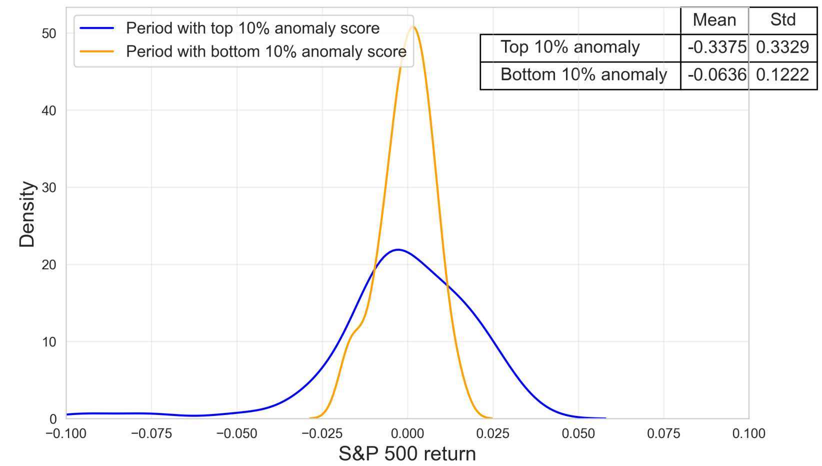


Analysis of anomaly score

Anomaly score & S&P500



Distribution of S&P500 Returns with Anomaly Scores



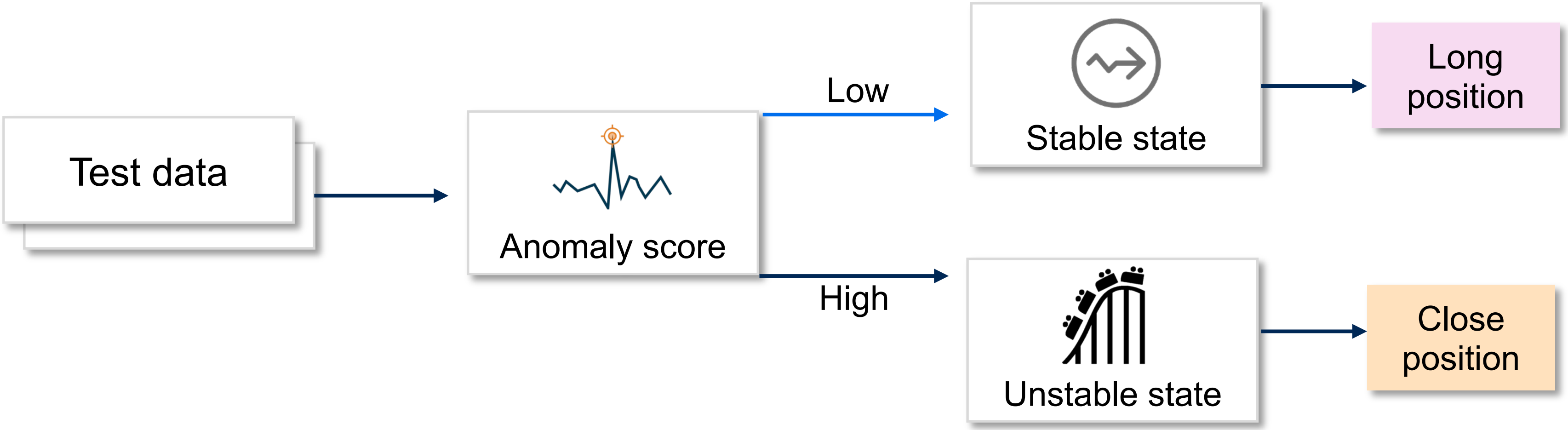
Empirical Analysis of Anomaly Scores

- The correlation coefficient between the anomaly score and S&P500 price data was **-0.5295**, indicating a quite strong inverse relationship.
- During the **bottom 10%** anomaly score days, the distribution of S&P500 returns are more concentrated around the center, indicating small market volatility.
- During the **top 10%** anomaly score days, the distribution of S&P500 returns exhibit much more left-skewed distribution.

Main idea obtained from anomaly score

Key findings from analysis of anomaly score

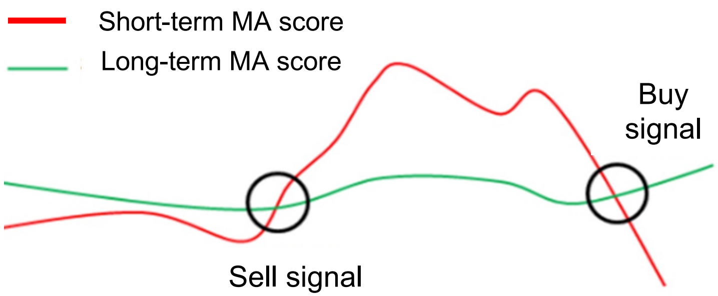
- Through the previous analysis on the anomaly score, we found that the stock price tends to fall and investment risk increases when the **anomaly score is high**.
- Therefore, we assume that the **stock price is unstable** when the **anomaly score is high**, and we can **buy** assets when the score is **low** and **sell** them when the score is **high** to prevent losses.



Anomaly Score as a Trading Signal

Moving average

- It is discovered that the price tends to move after a change in ano maly score, and a high anomaly score indicates market instability.
- To incorporate this insight, we calculated short-term and long-term moving averages to the anomaly score, and only bought when the short-term MA score exceeds the long-term MA score. We sold all holdings otherwise.



Voting sector signal strategy

1) Train GANs for each of the S&P 500 sector indices
2) Anomaly scores are derived for each of 11 sectors.
3) Aggregate 11 anomaly scores to derive the final signal
4) Final anomaly signal triggered when sector votes (anomaly signals) exceed thresholds obtained from
   - Equal-weighted voting
   - Market cap weighted voting

Evaluation Metrics

- Cumulative return
  $$C_R = \prod_{i=1}^{n}(1+R_i) - 1$$
- Annualized return
  $$A_R = (1 + C_R)^{1/c} - 1,$$ where c = number of years
- Annualized volatility $\sigma_R = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sigma_{R_i}^2} \times \sqrt{252}$
- Sharpe ratio $S_R = \frac{A_R - R_f}{\sigma_R}$,where R_f=risk-free rate

* In this study, R_f is set to 0 for simplicity.

- Maximum Drawdown (MDD)
  $$\text{MDD}(\%) = \frac{\textit{Peak value} - \textit{Lowest value}}{\textit{Peak value}} \times 100$$

Robust Portfolio Optimization via Anomaly Score

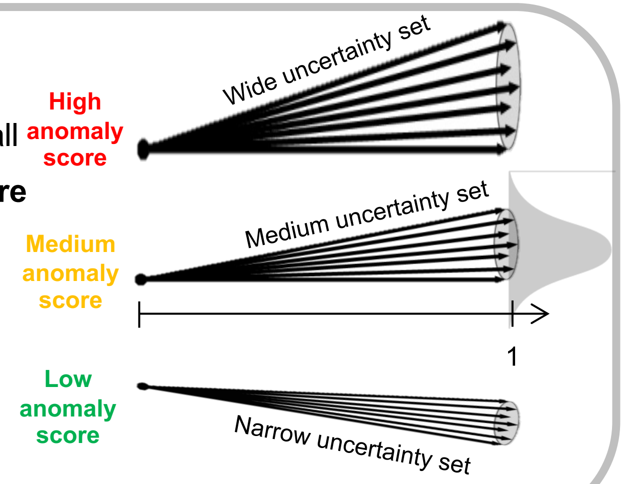
Idea

- From the previous experiment, we obtained anomaly scores for 11 sectors.
- We will utilize these scores to construct a robust sector portfolio by proportionally adjusting **the size of the uncertainty set based on the daily anomaly scores of each sector**.
- δ, which determines the uncertainty set size, is calculated as follows:

  where σ is standard deviation of asset and T is the total length of observed day

$$\delta = Z_{\frac{1+\delta_{pct}}{2}} \cdot \frac{\sigma}{\sqrt{T}}$$

- We adjust δ_{pct} to be low, high, or a medium value based on the anomaly score.



Baselines

$w \in \mathbb{R}^n$: portfolio weight
$\Sigma \in \mathbb{R}^{n \times n}$: covariance matrix
$\mu \in \mathbb{R}^n$: expected return
$\lambda \in \mathbb{R}$: risk-preference

1) S&P 500 Index
2) Mean-Variance Portfolio Optimization Model (MVO)

$$\text{minimize } w^T \Sigma w - \lambda w^T \mu$$
$$\text{subject to } w_1 + \cdots + w_n = 1$$

3) Vanilla Robust Portfolio Optimization Model (RO)

$$\text{minimize } w^T \Sigma w - \lambda \hat{\mu}^T w + \lambda \delta^T (w_+ + w_-)$$
$$\text{subject to } w_1 + \cdots + w_n = 1$$
$$w = w_+ - w_-$$
$$w_+ \geq 0, w_- \geq 0$$

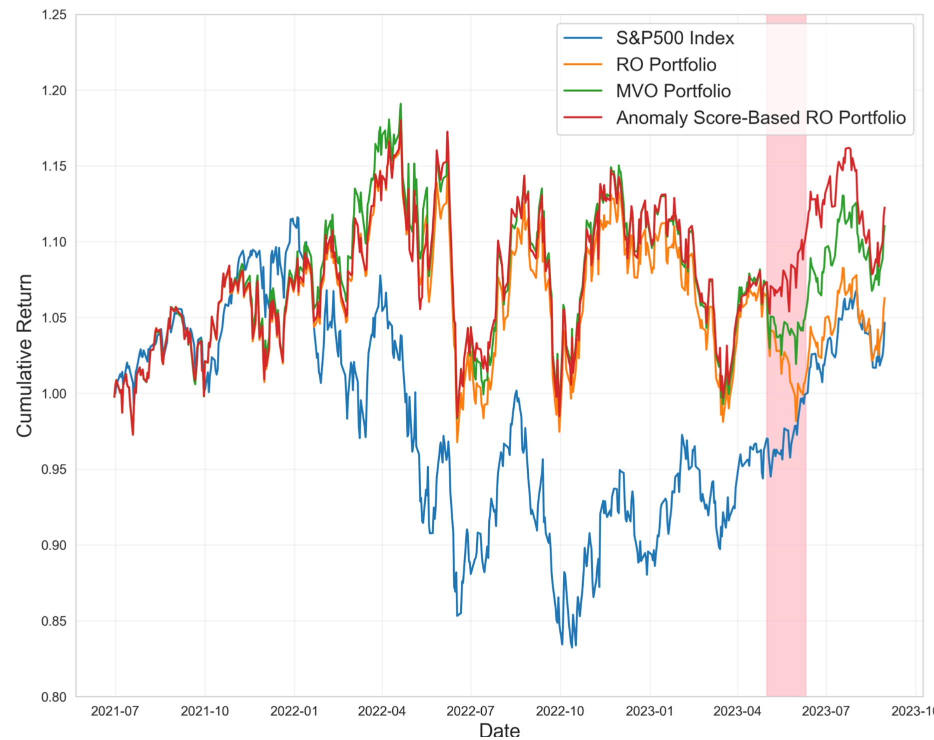
Anomaly Score-based Robust Portfolio Optimization

Portfolio construction rules

- The sum of asset weights equals 1, with **no single asset e xceeding 40%**.
- **Monthly rebalancing** conducted at the end of each month.
- Portfolio weight optimization performed for each model.
- Transaction cost set at 0.5% per rebalancing.

Evaluation results

	S&P500	MVO	RO	Anomaly score -based RO
Return	1.71%	**5.07%**	**2.97%**	**5.60%**
Volatility	19.07%	**18.57%**	**18.41%**	**18.44%**
SR	0.0897	**0.2731**	**0.1614**	**0.3037**
MDD	-25.43%	**-17.44%**	**-17.94%**	**-16.59%**



Parameter setting

- Risk aversion coefficient λ=0.5, Initial delta percent δ_{pct} =0.5,
- Mean-variance lookback period =12 months, Anomaly score lookback period = 3 months
- For **anomaly score-based robust optimization**:
  - Monthly, adjust δ_{pct} for each sector by checking how far they deviate from the average of the anomaly score's lookback period
  - Set δ_{pct} =0.1 if below **90%** of the average anomaly score
  - Set δ_{pct} = 0.9 if above **110%** of the average anomaly score

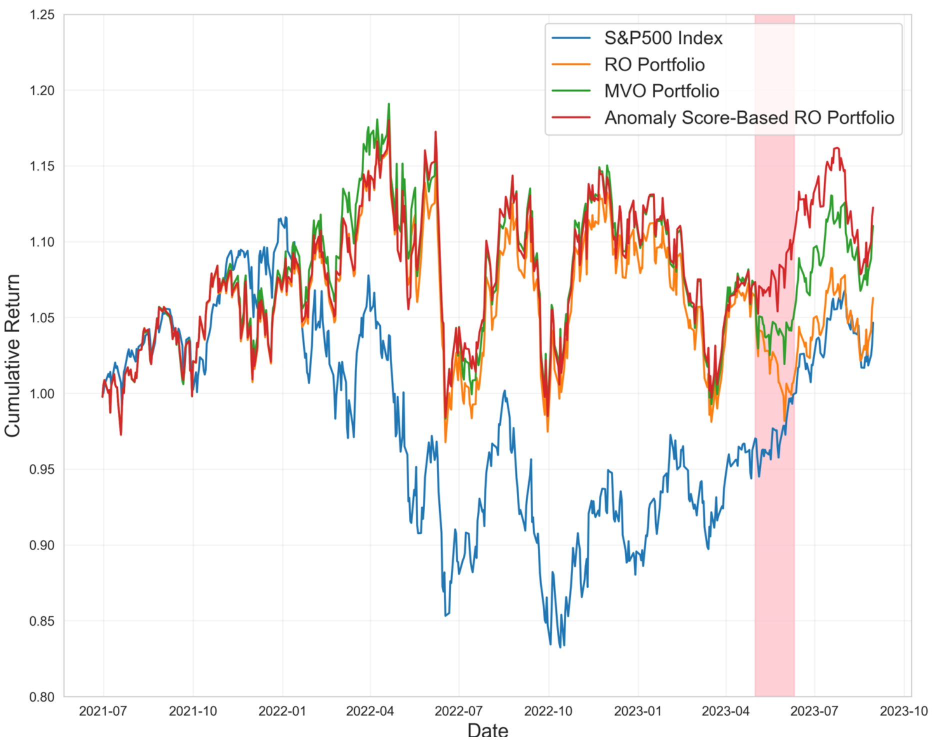
Anomaly Score-based Robust Portfolio Optimization

Portfolio construction rules

- The sum of asset weights equals 1, with **no single asset e xceeding 40%**.
- **Monthly rebalancing** conducted at the end of each month.
- Portfolio weight optimization performed for each model.
- Transaction cost set at 0.5% per rebalancing.

Evaluation results

	S&P500	MVO	RO	Anomaly score -based RO
Return	1.71%	**5.07%**	**2.97%**	**5.60%**
Volatility	19.07%	**18.57%**	**18.41%**	**18.44%**
SR	0.0897	**0.2731**	**0.1614**	**0.3037**
MDD	-25.43%	**-17.44%**	**-17.94%**	**-16.59%**



- **All portfolio optimization models** recorded higher final cumulative returns than the **S&P 500**, representing the market.
- **Anomaly score-based RO portfolio** achieved the **highest cumulative return, annualized return, and Sharpe ratio.**
- Notably, it had the **lowest MDD**, especially showing better loss mitigation from May to July 2023 compared to other models.
- Our method demonstrates superior performance in managing portfolio risk while outperforming the market index in terms of returns.

Weights for risk aversion coefficient λ

Parameter Setting and Results

- The parameter λ indicates the investor's risk preference: a high value signifies risk-taking, while a low value means risk aversion.
- Initial delta percent δ_{pct} =0.5 , Mean-variance lookback period =252, Anomaly score lookback period = 63
  - High threshold = average anomaly score * 1.1, Low threshold = average anomaly score * 0.9
- Results : Our method outperformed in terms of **return** and **Sharpe ratio** at λ=0.2, λ=0.5 and maintained the **lowest MDD** at λ=0.8.
  - Our method is more robust to different risk preferences than the traditional robust portfolio optimization.

λ	Benchmarks	Return	Volatility	Sharpe ratio	MDD
	S&P500	1.71%	19.07%	0.0897	-25.43%
Risk averse	MVO	**1.94%**	**17.65%**	0.1099	-18.06%
0.2	RO	1.04%	**17.39%**	0.0601	-18.10%
0.2	Anomaly score-based RO	**3.20%**	**17.46%**	**0.1832**	**-16.98%**
	MVO	**5.07%**	**18.57%**	0.2731	-17.44%
0.5	RO	**2.97%**	**18.41%**	0.1614	-17.94%
0.5	Anomaly score-based RO	**5.60%**	**18.44%**	**0.3037**	**-16.59%**
	MVO	**6.41%**	**18.76%**	**0.3417**	**-17.41%**
0.8	RO	**4.04%**	**18.61%**	**0.2171**	**-17.81%**
Risk taking 0.8	Anomaly score-based RO	**5.40%**	**18.66%**	**0.2892**	**-16.59%**

Conclusion and future work

- Used GANs to make robust portfolio optimization more flexible
  - Compared to static robust portfolio optimization, it could improve returns with minimal increment in risk

- The current work is done only for box uncertainty sets
  - We are extending this work for
    - Ellipsoidal uncertainty sets
      - Which is equivalent to covariance shrinkage estimators (Yin, Perchet, and Soupé, 2021)
    - Gerber statistics (Gerber, Markowitz, Ernst, Miao, Javid, & Sargen, 2022)
      - A simplified version of covariance matrix which can be helpful for mitigating sensitivity of MV optimization
- Also, we will test them on style portfolios as well
  - 6 portfolios on size & book-to-market (by Kenneth French)

IJCAI-24 Workshop on Recommender Systems in Finance (Fin-RecSys 24)

- **Date:** August 3-5, 2024 (TBD)
- **Paper submission deadline:** May 4, 2024
- **Author notification:** June 4, 2024
- **Location:** Jeju, Republic of Korea
- **Organizers**

Yongjae Lee

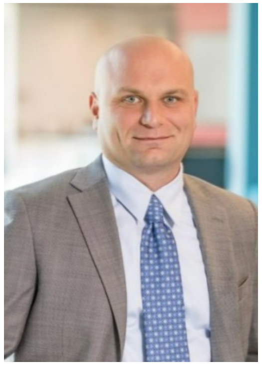
UNIST
(Lead Organizer)

John R.J. Thompson

University of
British Columbia

Dhagash Mehta

BlackRock



Thomas De Luca

Vanguard

Richard Mccreadie

University of Glasgow

Jaesik Choi

KAIST

Min Hee Kim

Hana Institute
of Technology

https://sites.google.com/view/fin-recsys2024/

Advertisement

- **Guest editors**

  Matt Davison (Western University, Canada)
  Yongjae Lee (Ulsan National Institute of Science and Technology)
  Dhagash Mehta (BlackRock, Inc.)
  Alberto G. Rossi (Georgetown University)
  John R.J. Thompson (University of British Columbia, Canada)

https://think.taylorandfrancis.com/special_issues/statistical-machine-learning-investor-modelling/

Advertisement

ICAIF'24

5th ACM International Conference on AI in Finance

General Chairs

- Dhagash Mehta (BlackRock)
- Guiling "Grace" Wang (New Jersey Institute of Technology)

Program Chairs

- Senthil Kumar (Capital One)
- Hao Ni (UCL)

https://ai-finance.org/

Workshop Chairs

- Bo An (Nanyang Technical University)
- Yongjae Lee (Ulsan National Institute of Science and Technology)
- Zhen Zeng (JP Morgan)

Thank you for listening!