

근위 정책 최적화를 활용한 자산 배분에 관한 연구

A Study on Asset Allocation Using Proximal Policy Optimization

이우식*

Lee Woo Sik*

〈Abstract〉

Recently, deep reinforcement learning has been applied to a variety of industries, such as games, robotics, autonomous vehicles, and data cooling systems. An algorithm called reinforcement learning allows for automated asset allocation without the requirement for ongoing monitoring. It is free to choose its own policies. The purpose of this paper is to carry out an empirical analysis of the performance of asset allocation strategies. Among the strategies considered were the conventional Mean-Variance Optimization (MVO) and the Proximal Policy Optimization (PPO). According to the findings, the PPO outperformed both its benchmark index and the MVO. This paper demonstrates how dynamic asset allocation can benefit from the development of a reinforcement learning algorithm.

Keywords : Quantitative Finance, Business Analytics, FinTech, Robo-Advisor, Reinforcement Learning

* 정회원, 제1저자, 경상국립대학교 경영대학, 교수
E-mail: woosiklee@gnu.ac.kr

* College of Business Administration, Gyeongsang National University

1. 서론

투자이론에서 위험을 감소시키기 위해 상호 간의 상관관계 수준이 낮은 자산들을 묶어 투자하는 행위를 ‘분산 투자’라고 한다[1]. 투자이론에 있어 분산 투자는 이론적 기초이며, 이 개념을 토대로 ‘현대 포트폴리오 이론’이 구체화 및 체계화되었다 [1]. 가장 대표적인 현대 포트폴리오 이론은 1952년 해리 마코위츠(Harry M. Markowitz)의 평균-분산 최적화 모형이고, 이 이론에서는 분산 투자로 구성된 포트폴리오가 위험 감소에 긍정적 기여를 한다고 밝히고 있다[1][2]. 평균-분산 최적화 모형은 최적의 자산 배분을 구축하는 이론적 토대로써 학계 및 실무에서 많이 활용된다 [2].

현재도 자산운용사, 투자은행, 연기금, 증권회사 등 대형 기관을 포함한 많은 투자자들은 최적의 자산 배분을 구축해 나가는 과정에서 평균-분산 최적화 모형을 활용하고 있다 [2]. 하지만 평균-분산 최적화에 입각한 자산 배분은 수학적으로 제공된 매우 우수함에도 불구하고, 여러 현실적 한계들이 지적되고 있다. 평균-분산 최적화 모형의 한계 중 하나는 자산 배분이 자산들의 기대수익률과 공분산에 매우 민감하기 때문에 특정 자산에 투자 비중이 집중되는 상황을 초래하여 결과적으로 외표본에서 분산 효과가 저하되는 문제가 발생할 소지가 있다 [3]. 이처럼 특정 자산에 집중되는 포트폴리오는 실무적으로 받아들이기 쉽지 않다 [2]. 평균-분산 최적화 모형의 또 다른 문제점은 자산군 수익률의 정규성을 가정하여 위험을 과소평가할 수 있다는 점이다 [2]. 실증적으로 두꺼운 꼬리 등 자산군 수익률의 비정규성을 의심하게 하는 상황이 목격된다 [2]. 이 외에도 평균-분산 최적화 모형은 자산들의 수익률 분포가 일정하게 유지될 것을 가정하고 있다 [2]. 그러나 실제로 과거의 투자 수익률과 변동성이 미래에 대폭 바뀔 수 있다.

Michael와 Grauer[4]는 실증 분석을 통해 포트폴리오의 추정 수익률 평균과 공분산 행렬이 기대한 것과 달리 추정 오류를 범하게 되었을 때는 포트폴리오의 성능이 저하된다는 것을 확인했다.

이처럼 평균-분산 최적화 모형의 실제적 한계가 지속되면서, 평균-분산 최적화 모형의 제한을 극복하기 위한 대안으로 최적화 과정에 많은 제약과 통제를 부가하기도 하지만, 금융시장 상황이나 경제 상황의 변화에 따른 동적 최적 자산 배분의 필요성과 도입도 주장하고 있다. 최명훈과 성주호 [5]는 라이프 스타일을 반영한 동적 자산 배분을 정적 자산 배분과 비교하여 연금 자산 운용에 동적 자산의 당위성을 제시하였고, 최인우와 하규수 [6]은 여러 요인에 적용하는 가중치를 동적으로 전환시켜 얻은 포트폴리오가 유의하고 안정적인 초과수익을 얻을 수 있는 것으로 평가했다. Harman 외[7]는 정적 자산 배분과 비교해 동적 자산 배분은 시장 상황에 따라 단기간의 기회를 체계적으로 파악하여 성과를 추가로 올릴 수 있게 해 주어 동적 자산 배분이 정적 자산 배분을 보완하는 개념으로 제시하였다.

더불어 정보이론, 통계 그리고 기계학습 기법을 활용한 입력변수 예측 등 마코위츠 모형의 약점을 보완하고자 하는 여러 연구가 이루어지고 있다. Bouchard 등[8]은 엔트로피를 바탕으로 투자 종목에 관한 정보가 부족할 때, 소수의 주식에 투자가 집중되는 현상을 완화하고자 하였다. 김영현[9]는 추적 신호를 적용하여 마코위츠 모형의 종목 선정 능력을 향상시켰고, Fan 외[10]은 팩터 모형에 주성분 분석을 통해 표본 공분산을 분해하여 추정하였다. Pantaleo 등[11]은 고차원 자료에 적용하기 쉽도록 목표 공분산까지 표본 공분산을 수축하는 기법을 사용하여 추정하였다. 다음으로 Cai 등[12]는 역공분산 행렬에 L_1 제약을 두어 정밀도 행렬을 추정하였다. Millington과 Nirangjan[13]은 그래프 라

쓰를 사용하여 희소한 정밀도 행렬을 추정하였다.

이처럼 여러 기법과 새로운 실험을 시도함으로써 자산 배분의 다양성 증진과 기존 자산 배분의 한계 보완 등에 기여하고 있다. 최근 강화학습의 출현과 폭발적 성장을 통해 자산 배분 관련 국내 연구들이 수행되고 있지만, 여전히 미흡한 수준에서 벗어나지 못하고 있다. 기존 연구 중 김주봉 등[14]은 A3C-DPG기반의 암호화폐 포트폴리오를 제시하였고, 윤성식과 이석준[15]도 A3C 알고리즘을 이용하여 개별 주식 종목에 대한 학습보다 주식시장에 대한 학습이 더 우수한 성과를 창출할 수 있다고 제시하였다. 하지만 이들 연구들은 암호화폐와 개별 주식에 미치는 A3C 강화학습 알고리즘 영향에 대해 분석하였다. 이에 반해 본 연구는 가장 범용적으로 사용되는 평균-분산 최적화를 준거 자산 배분 모형으로 보고, 위험조정 수익 극대화 측면에서 자산 가격의 변동 등 시장 상황을 반영할 수 있는 근위 정책 최적화를 통해 자산 배분을 개선할 수 있는지 살펴보고자 하였다.

본 논문은 다음과 같은 구성 체계로 기술되어 있다. 제1절 서론에서는 연구의 배경과 필요성을 밝힌다. 제2절에서는 주요 방법론인 평균-분산 최적화와 근위 정책 최적화를 활용한 자산 배분에 대한 설명을 소개하였으며, 제3절에서는 실증 분석 및 그 결과를 확인한다. 마지막으로 제4절에서는 결론과 시사점을 제시한다.

2. 이론적 배경

2.1 평균-분산 최적화

해리 마코위츠의 현대 포트폴리오 이론은 가능한 여러 투자 기회들 중에서 최상의 위험-수익률 조합을 찾아내어 최적의 투자 기회를 선택·결정함

으로써 분산 투자의 기본 논리로서 광범위하게 활용되고 있다[16]. 마코위츠가 제안한 평균-분산 최적화 모형은 자산의 기대수익률과 공분산으로 포트폴리오의 최적 구성 방법을 제시하고 있다.

$$\text{Min } \sigma_p^2 = \sum_{i=1}^N \sum_{j=1}^N \sigma_{ij} w_i w_j \quad (1)$$

$$\text{subject to } \sum_{j=1}^N w_j \mu_j = c$$

$$\sum_{j=1}^N w_j = 1 \quad w \geq 0 \text{ for } j = 1, \dots, N [17]$$

- N : 자산의 수
- w_j : 자산 j 의 투자비중 ($j = 1, 2, \dots, N$)
- μ_j : 자산 j 의 기대수익률 ($j = 1, 2, \dots, N$)
- σ_{jj} : 자산 j 의 기대수익률에 대한 분산 ($j = 1, 2, \dots, N$)
- σ_{ij} : 자산 i 의 기대수익률과 자산 j 의 기대수익률의 공분산
- c : 포트폴리오의 목표기대수익률
- σ_p^2 : 포트폴리오의 수익률에 대한 분산

2.2 근위 정책 최적화

강화학습은 $\{ S, A, P, R, \gamma \}$ 로 이루어진 마르코프 결정 과정으로 정의되는 환경으로부터 누적 보상값의 기대값을 최대화하는 최적화 기법이다 [18]. 여기서 S 는 유한의 크기를 취하는 상태 집합이고, A 는 유한의 크기를 취하는 행동 집합이며, $P(s'|s, a)$ 는 상태 전이에 대한 확률로서 현재의 상태인 $s \in S$ 에서 행동 $a \in A$ 를 하였을 때 다음 상태가 $s' \in S$ 이 되도록 하는 확률 분포를 의미한다. 또한 R 은 보상 함수, $\gamma \in (0, 1)$ 은 할인 계수를 나타낸다[18].

강화학습의 학습 순서는 각각의 시간 단계별로 정의된 강화학습 에이전트가 주어진 환경 조건에서 현재의 상태를 주시하여 이를 기반으로 행동을 선

택하고, 이때 환경 상태에 변화가 생겨나면서 정의된 에이전트는 행동에 의한 보상을 받는다. 학습 초기에 에이전트는 무작위 행동을 하지만, 학습이 점차 진행되면서 보상의 수준을 더 높일 수 있는 행동으로 선화하여 학습을 행하게 된다. 강화학습은 초기 상태로부터 정책을 기반으로 한 연속적 행동을 취하게 되었을 때 획득하는 기대누적보상을 최대화하는 정책을 발견하는 것을 목적으로 하는데, 이것을 최적 정책이라고 지칭한다[18].

심층강화학습은 마르코프 결정 과정을 사용하여 순차적 의사결정 문제를 모형화하고, 가치 함수나 정책 함수의 근사자로 인공신경망을 강화학습에 활용하여 문제를 해결하는 방법론이다. 심층강화학습 기법들 중 가치 함수 및 정책 함수의 근사를 모두 활용한 근위 정책 최적화(Proximal Policy Optimization)는 비활성 정책 학습법 기반으로 목표 신경망을 갖고, 액터 신경망이 행동을 연산하고 크리틱 신경망이 행동 가치에 대한 계산을 통해 행동 개선을 도모하는 알고리즘이다[19].

근위 정책 최적화 알고리즘은 다양한 강화학습 기법들 중에서 정책 경사 기법에 해당된다[20]. 정책 경사 기법은 이와 구분되는 가치 경사 기법과 비교할 때, 행동이 점진적으로 변화하여 수렴성이 상대적으로 크다는 장점을 갖고 있다[20]. 이와 더불어 연속 행동 공간에 대해 효율적이면서도 확률론적인 정책을 효과적으로 학습할 수 있다는 장점도 있다[20]. 대체로 정책 경사도는 정책 파라미터(θ)에 대해 다음과 같은 목적 함수 $L^{PolicyGradient}$ 를 최대화할 수 있는 방향으로 계산이 이루어진다[20].

$$L^{PolicyGradient} = \hat{E}[\log \pi_{\theta}(a_t | s_t) \hat{A}_t] \quad (2)$$

식 (2)에서 a_t 와 s_t 는 시간 t 에서 수행한 각각의 행동 및 상태를 뜻한다. 또한 π_{θ} 는 확률론적인 정

책을 뜻하고, \hat{A}_t 는 이득 함수(Advantage Function)의 추정값을 뜻한다[20]. \hat{E}_t 는 샘플링된 데이터에서의 평균값을 뜻한다[20]. $L^{PolicyGradient}$ 는 가치 경사와 비교할 때 점진적 정책의 변화를 유도하게 된다. 그렇지만 파라미터 공간상에서 이루어지는 점진적 변화가 여전히 정책 공간상에서는 매우 큰 변화를 유발하게 될 수도 있다는 한계점을 지닌다[20].

근위 정책 최적화는 정책 공간상에서 발생하는 변화를 감안할 목적으로 식 (3)과 같이 현재 정책 확률과 이전 정책 확률 간의 비율(r_t)을 정의한다. 또한 식 (4)와 같이 이를 활용한 대체 목적 함수(Surrogate Object Function)를 구성한다[20].

$$r_t(\theta) = \frac{\pi_{\theta}(a_t | s_t)}{\pi_{\theta_{old}}(a_t | s_t)} \quad (3)$$

$$L^{SOFF}(\theta) = \hat{E}_t[r_t(\theta) \hat{A}_t] \quad (4)$$

$$L^{CLIPPING}(\theta) = \hat{E}[\min(r_t(\theta) \hat{A}_t, clip(r_t, 1 - \epsilon, 1 + \epsilon) \hat{A}_t)] \quad (5)$$

그다음 필요 이상으로 발생할 수 있는 정책 갱신의 문제를 방지하고자 대체 목적 함수인 L^{SOFF} 에 대해 클리핑 기법을 활용해 도출한 새로운 목적 함수인 $L^{CLIPPING}$ 을 사용한다. 여기에서 ϵ 는 초매개변수를 뜻한다[20].

이득 함수인 \hat{A}_t 는 식 (6)에서 보듯이 상태에 대한 가치 함수인 $V(s)$ 를 통해서 추정이 가능하다[20].

$$\hat{A}_t = -V(s_t) + v_t + \gamma v_{t+1} + \dots + \gamma^{T-t+1} v_{T-1} + \gamma^{T-t} V(s_T) \quad (6)$$

식 (6)에서 T 는 총 시간의 스텝 수이고, t 는 $[0, T]$ 범위에서의 시간 인덱스이며, s_t 는 t 에서의 상태이다. 또한 v_t 는 t 에서의 보상이고, γ 는 할인

요소를 나타낸 것이다[20]. 신경망을 이용해 학습을 행하는 모델의 경우 가치 함수도 다음 식 (7)과 같이 목적 함수 L^{Value} 를 통해서 가치 파라미터를 갱신하게 된다[20].

$$L^{Value}(\mu) = \hat{E}_t [(V_\mu(s_t) - V_t)^2] \quad (7)$$

식 (7)에서 V_μ 는 μ 로 구성된 가치 함수의 신경망을 뜻하며, $V_t = v_{t+1} + \gamma V_\mu(s_{t+1})$ 이다[20].

3. 실증 분석

3.1 자료의 구성

본 연구에서 활용할 표본은 미국 다우존스 산업 평균 인덱스로 전체 미국 시장을 대표할 수 있는 지수는 아니지만, 인덱스를 구성하는 주식 수가 우량기업 30개로 상대적으로 쉽게 추종 인덱스 움직임에 따른 포트폴리오를 구성할 수 있다. 또한 포트폴리오의 자산을 희소하게 구성할수록 관리경비가 낮아지고, 구성 주식의 가중치 변화가 적도록 안정적으로 구성하면 거래비용이 낮아질 수 있어 이를 통해 투자 포트폴리오의 다변화를 원하는 국내 투자자가 미국 금융시장에 투자할 수 있는 간단한 투자 수단에 부합한다고 볼 수 있다. 실험을 위해 2011년 1월 3일부터 2020년 12월 30일까지 일별 종가자료를 활용하고, 모형의 성과 측정을 위해 2021년 동안의 투자 기간 자료를 확보하였다.

다우존스 산업 평균 인덱스 일별 종가에 대한 기술통계량(평균, 표준편차, 왜도와 첨도)은 Table 1에서 살펴볼 수 있다. 본 인덱스에 나타나는 음의 값을 취하는 첨도는 꼬리가 정규분포의 폭보다 좁은 것을 나타내고, 인덱스 수익률의 왜도가 음의 값을 취하는 것은 부정적 극단 현상의 발생 가

Table. 1 Descriptive Statistics

	DJI	DJI Return(%)
Mean	20577.68	0.000470
Median	18134.21	0.000655
Max.	36488.63	0.113650
Min.	10655.30	-0.129265
S.D.	6637.568	0.010786
Skewness	0.571073	-0.617947
Kurtosis	-0.623375	21.679325

능성이 정규분포에 비해 높다는 것을 뜻한다[18].

$$\text{주가지수 변화율} = \ln(\text{주가지수}(t) / \text{주가지수}(t-1)) \quad (8)$$

3.2 모형의 추정 및 분석

평균-분산 최적화와 근위 정책 최적화를 활용한 자산 배분을 통한 포트폴리오의 성과 비교 및 검증으로 강화학습 에이전트가 지속해서 금융시장의 변화를 올바르게 감지하고 최적의 자산 배분을 할 수 있는지를 확인하였다. 즉, 평균-분산과 근위 정책 최적화를 활용한 자산 배분을 위해 미국 다우 인덱스를 구성하고 있는 종목의 일별 수익률과 이에 대한 상관행렬을 상태변수로 사용하고 투자 자산 비중에 따른 샤프지수의 비교와 검증을 시행하였다. 근위 정책 최적화를 활용한 자산 배분의 경우, 다층 퍼셉트론을 정책 신경망에 256개의 유닛 수를 가진 액터와 크리티크 신경망 공통 아키텍처는 2개로 설정하였다. 더불어 배치 크기는 128개, 학습률은 0.0001로 구성하였다. 공매도와 증권거래세 등의 거래비용은 고려하지 않았다.

정책 경사 알고리즘으로 에이전트를 학습시킬 때 고려할 점은 에이전트의 성능이 급히 저하되는 성능 붕괴 현상이 발생할 수 있다는 것이다. 이런 상황이 되면 에이전트는 좋지 않은 성능 궤적을 만들

것이고, 그 성능 궤적을 이용하여 정책을 계속 교유할 것이기 때문에 원래 상태로 회복하기 어려워진다. 더불어 활성 정책 알고리즘이 자료를 재사용하지 못하기 때문에 표본 비효율이 된다[21].

이에 본 논문에서는 강화학습 에이전트가 지속적인 학습 과정에서 오류를 감소시키고 투자 손실을 최소화하는 방향으로 자산 배분을 수행할 수 있는지 기존의 평균-분산 최적화를 활용한 자산 배분과 비교하는 연구를 수행했다. 그 결과 Table 2에서 보듯이 모든 자산 배분에서 정(+)의 샤프지수를 보였는데, 이것은 위험 대비 투자수익이 발생했음을 의미한다. 특히, 근위 정책 최적화가 클리핑(Clipping) 기법을 적용해 필요 이상의 정책 갱신을 방지할 뿐만 아니라, 액터-크리틱모형 기반으로 학습하기 때문에 동작 확률 분포에 따라 동작 선택 후 보상을 받고, 이것을 상태 가치와의 비교를 통해서 이익을 계산하기 때문에 최적의 정책을 학습시킬 확률이 높아졌다고 판단된다[22]. 샤프지수가 가장 높을 때는 근위 정책 최적화를

활용한 자산 배분을 한 경우(1.66)로 나타났고, 평균-분산 최적화에서는 인덱스보다 훨씬 낮은 위험 대비 높은 투자 성과(0.77)를 보여 주었다. 또한 근위 정책 최적화의 최대자본인하율(-0.06)이 다소 높지만, 이는 높은 수익률 증가로 인해 샤프지수가 인덱스와 평균-분산 최적화를 활용한 자산 배분보다 증가했음을 알 수 있다.

마지막으로, Fig. 1에서와 같이 평균-분산 최적화와 근위 정책 최적화를 활용한 자산 배분의 누적 수익률을 살펴보면, 강화학습 알고리즘을 활용한 자산 배분에서 상대적으로 더 높은 성과를 보여 주었다.

4. 결론

하나은행에 따르면 국내 로보어드바이저 시장 규모가 2025년에는 30조 원으로 성장할 것으로 내다보고 있고, 코스콤에 따르면 작년 11월 기준 국내 로보어드바이저 가입자는 총 41만여 명으로 로보어드바이저에 많은 관심을 보이고 있다 [23][24]. 이러한 상황 속에서 인공지능을 포함한 기계학습 기반의 자산관리 서비스 제공을 위한 신생 기업들이 우후죽순 생겨나고 있다. 주요 금융업체들은 인공지능과 빅데이터 기반의 금융 서비스 고도화를 위한 업무 협약이 활발하게 진행되고 있다. 이와 더불어 대한민국 금융위원회에서도 4차 산업혁명기술과 함께 안정적·균형적 발전을 도모하기 위해 핀테크를 더욱 활성화하는 내용의 「전자금융거래법」 개정을 추진하고 있다. 이처럼 4차 산업혁명기술 기반의 자산관리에 대한 중요성이 커지면서 향후 계량금융의 성장성은 매우 높을 것으로 예상된다[18].

본 연구에서는 자산 배분에 가장 범용적으로 사용되는 해리 마코위츠의 평균-분산 최적화와 근

Table 2. Performance of Asset Allocation

	Return	Volatility	Sharpe Ratio	MDD
Index	0.2090	0.1242	1.5966	-0.0661
MVO	0.0763	0.1022	0.7712	-0.0661
PPO	0.2120	0.1202	1.6604	-0.0687



Fig. 1 Cumulative Returns of MVO and PPO

위 정책 최적화를 활용한 자산 배분의 성과를 비교·분석함으로써 학술적 가치가 있다고 할 수 있다. 즉, 마코위츠는 현대 포트폴리오 이론을 창안하고 적절한 포트폴리오 가중치를 선택해 분산 투자를 최적화하는 수학적 도구를 제공했다. 하지만 마코위츠의 평균-분산 최적화는 표본 내, 즉 후향적인 최적화를 보여 주고 있다. 그러나 실제로 자산 배분 최적화를 위해서는 전향적인 입력과 출력이 필요하다. 이에 강화학습 에이전트가 지속적인 학습 과정을 통해 위험조정 수익 극대화 측면에서 자산 배분을 수행할 수 있는지 기존의 평균-분산 최적화를 활용한 자산 배분과 비교·분석하였다. 본 연구의 주요 분석 결과는 다음과 같다.

첫째, 평균-분산 최적화와 근위 정책 최적화 기반의 자산 배분의 위험 대비 높은 투자 성능을 비교해 보면, 평균-분산 최적화 모형에서 아주 낮은 성과를 보여 주었다. 이러한 결과는 특히 동적 자산 배분의 중요성을 함축하는 결과라 할 수 있다. 즉, 강화학습 에이전트는 시장 환경의 변화를 미리 감지하고 이에 신속히 대응해 시장의 위험 구조 및 수익률 간의 최적 균형을 도출해 내는 것이다. 이로 인해 강화학습 에이전트는 수익 기회를 온전히 유지하면서도 변동성은 감소시키고 갑작스런 손실 또한 방지할 수 있는 것이다.

둘째, 근위 정책 최적화를 활용한 자산 배분의 샵프지수가 인덱스와 평균-분산 최적화 보다 더 높은 수치를 기록했다. 이는 근위 정책 최적화가 액터-크리틱 모형 기반으로 학습하기 때문에 동작 확률 분포로 동작 선택 후 보상을 받고, 이것을 상태 가치와 비교를 하여 이익을 계산하기 때문에 최적의 정책을 학습시킬 확률이 높아졌다고 판단된다[25].

강화학습을 활용한 자산 배분의 유용성에 대해서 인정한다고 하더라도 실제로 자산 배분을 운용하고, 성과를 내기는 쉽지 않은 일이다. 강화학습

에 필요한 경제·금융 환경, 정책 그리고 세부적인 구현 설정과 조합 등의 필요성으로 인해 자산 배분에서의 응용에 어려움이 존재하고, 이를 적절히 최적화하지 못하면 투자 성과를 이끌어 낼 수 없다. 강화학습 에이전트가 지속적으로 경제 상태와 금융시장의 변화를 올바르게 감지하고, 투자 수익률에 대한 기대치를 정확하게 갱신할 수 있을 때만이 제 역할을 할 수 있을 것이다.

본 논문에도 향후 몇 가지 보완할 점이 필요하다. 자산 배분 자체가 가지고 있는 특성 이외에도 해당 최적화를 자산 배분 운용에 실제 적용하는 과정에서 다양한 제약조건이 존재한다[26]. 그 중에서 최적 투자 비중 추종을 위해 수행하는 리밸런싱의 횟수에 따라 거래비용 등의 제약이 따르게 된다[26]. 이에 실제 자산 운용에 있어 공매도, 증권거래세 그리고 매매 주체에 따른 거래비용을 고려한 연구가 필요하다.

참고문헌

- [1] D. Ahn, and S. Park, "Linear programming models using a Dantzig type risk for portfolio optimization," *The Korean Journal of applied Statistics*, vol. 35, no. 2, pp.229-250, (2022).
- [2] W. Yoo, and Y. Choi, "A Study on the Improvement of Strategic Asset Allocation Using Global Investor's Reference Portfolio," *Korea Finance Association Conference*, pp. 214-324, (2019).
- [3] J. Hahn, S. Park, and H. E. Young, "Evaluating the Empirical Performance of Risk-based Portfolio Strategies in the Korean Stock Market," *The Korean Journal of applied Statistics*, vol. 45, no. 2, pp.247-284, (2016).
- [4] B. Michael, and R. Grauer, "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and

- Computational Results,” *Review of Financial Studies*, vol. 4, pp.315-342, (1991).
- [5] M. h. Choi, and J. Sung, “Why should Government Alleviate the Current Regulation of Defined Contribution Pension Asset Management in Korea?,” *Journal of The Korean Data Analysis Society*, vol. 13, no. 5, pp.2629-2642, (2011).
- [6] I. Choi, and K. Ha, “A Study on a Method for Portfolio Construction using Dynamic Multi-Factor Model and Black-Litterman-Herold Model,” *Journal of The Korean Data Analysis Society*, vol. 13, no. 5, pp.2599-2613, (2011).
- [7] A. Harman, v. d. L. Epcó, S. Kej, and K. Petr, “Dynamic Asset Allocation,” *Colonial First State Global Asset Management Multi-Asset Solutions Research Papers*, no. 7, New York, (2017).
- [8] J. Bouchaud, M. Potters, and J. Aguilar, “Missing Information and Asset Allocation,” *Science & Finance (CFM) working paper archive 500045*, Science & Finance, Capital Fund Management, (1997).
- [9] Y. Kim, H. Kim, and S. Kim, “Application of Tracking Signal to the Markowitz Portfolio Selection Model to Improve Stock Selection Ability by Overcoming Estimation Error,” *Journal of the Korean Operations Research and Management Science Society*, vol. 41, no. 3, pp.1-21, (2016).
- [10] W. J. Fan, Y. Liao, and M. Mincheva, “Large Covariance Estimation by Thresholding Principal Orthogonal Complements,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 75, no. 4, pp.603-680, (2013).
- [11] E. Pantaleo, M. Tumminello, F. Lillo, and R. Mantegna, “When do improved covariance matrix estimators enhance portfolio optimization? An empirical comparative study of nine estimators,” *Quantitative Finance*, vol. 11, pp. 1067-1080, (2011).
- [12] T. T. Cai, W. Liu, and X. Luo, “A Constrained L1 Minimization Approach to Sparse Precision Matrix Estimation,” *Journal of the American Statistical Association*, vol. 106, pp.591-607, (2011).
- [13] T. Millington, and M. Niranján, “Robust Portfolio Risk Minimization Using the Graphical Lasso,” *International Conference on Neural Information Processing*, pp.863-872, (2019).
- [14] J. Kim, J. Heo, H. Lim, D. Kwon, and Y. Han, “Blockchain Based Financial Portfolio Management Using A3C,” *KIPS Transactions on Computer and Communication Systems*, vol. 8, no. 1, pp.17-28, (2019).
- [15] S. Yoon, and S. Lee, “Stock Portfolio Construction and Trading Simulation Using Reinforcement Learning,” *Management & Information Systems Review*, vol. 40, no. 4, pp.185-203, (2021).
- [16] K. Cho, S. Lee, and J. Kim, “An Empirical Study on the Risk Diversification Effect of REITs,” *Korean Journal of Construction Engineering and Management*, vol. 14, no. 1, pp.23-31, (2013).
- [17] I. Song, “Using Decision Making Model for Asset Allocation,” *Asset Management Review*, vol. 7, no. 2, pp.46-64, (2019).
- [18] W. Lee, “Performance Evaluation of Portfolio using a Deep Q-Networks,” *Journal of Next-generation Convergence Information Services Technology*, vol. 10, no. 4, pp.459-470, (2021).
- [19] D. Lee, and M. Kwon, “Combating Stop-and-Go Wave Problem at a Ring Road Using Deep Reinforcement Learning Based Autonomous Vehicles,” *The Journal of Korean Institute of Communications and Information Sciences*, vol. 46, no. 10, pp.1667-1682, (2021).
- [20] H. Chae, D. Lee, S. Park, H. Choi and H. Park, “Time-varying Proportional Navigation Guidance using Deep Reinforcement Learning,” *Journal of the Korea Institute of Military Science and Technology*, vol. 23, no. 4, pp. 399-406, (2020).
- [21] Graesser, L. and Keng, W. L. : *Foundations of*

- Deep Reinforcement Learning: Theory and Practice in Python.1st Ed., Addison-Wesley Professional. Press, Boston, (2019).
- [22] W. S. Lee, “A Study on the Portfolio Performance Evaluation using Actor-Critic Reinforcement Learning Algorithms,” Journal of The Korean Society of Industry Convergence, vol. 25, no. 3, pp.467-476, (2022).
- [23] Koscom Newsroom, “<https://newsroom.koscom.co.kr/27881>”
- [24] Hankyung Magazine, “<https://magazine.hankyung.com/business/article/202201058897b>”
- [25] W. S. Lee, “A Study on Portfolio based on a Deep Deterministic Policy Gradient,” Journal of Next-generation Convergence Information Services Technology, vol. 11, no. 3, pp.287-298, (2022).
- [26] J. Lee, Y. Lee, and Y. Cho, “Performance analysis of sector index portfolios using the GOP model,” Journal of the Korean Data Analysis Society, vol. 24, no. 2, pp.823-841, (2022).

(접수: 2022.07.11. 수정: 2022.08.07. 게재확장: 2022.08.10.)