ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2020 March: 26(1): 135~149 http://dx.doi.org/10.13088/jiis.2020.26.1.135

XGBoost를 활용한 리스크패리티 자산배분 모형에 관한 연구

김영훈
국민대학교
비즈니스IT 전문대학원
(*dudgnskr92@naver.com*)

최흥식
국민대학교
비즈니스IT 전문대학원
(*hschoi@kookmin.ac.kr*)

김선웅
국민대학교
비즈니스IT 전문대학원
(*swkim@kookmin.ac.kr*)

인공지능을 기반으로 한 다양한 연구들이 현대사회에 많은 변화를 불러일으키고 있다. 금융시장 역시 예외는 아니다. 로보어드바이저 개발이 활발하게 진행되고 있으며 전통적 방식의 단점을 보완하고 사람이 분석하기 어려운 부분을 대체하고 있다. 로보어드바이저는 인공지능 알고리즘으로 자동화된 투자 결정을 내려 다양한 자산배분 모형과 함께 활용되고 있다. 자산배분 모형 중 리스크패리티는 대표적인 위험 기반 자산배분 모형의 하나로 큰 자산을 운용하는 데 있어 안정성을 나타내고 현업에서 역시 널리 쓰이고 있다. 그리고 XGBoost 모형은 병렬화된 트리 부스팅 기법으로 제한된 메모리 환경에서도 수십억 가지의 예제로 확장이 가능할 뿐만 아니라 기존의 부스팅에 비해 학습속도가 매우 빨라 많은 분야에서 널리 활용되고 있다. 이에 본 연구에서 리스크패리티와 XGBoost를 장점을 결합한 모형을 제안하고자 한다. 기존에 널리 사용되는 최적화 자산배분 모형은 과거 데이터를 기반으로 투자 비중을 추정하기 때문에 과거와 실투자 기간 사이의 추정 오차가 발생하게 된다. 최적화 자산배분 모형은 추정 오차로 인해 포트폴리오 성과에서 악영향을 받게 된다. 본 연구는 XGBoost를 통해 실투자 기간의 변동성을 예측하여 최적화 자산배분 모형의 추정 오차를 줄여 모형의 안정성과 포트폴리오 성과를 개선하고자 한다.

본 연구에서 제시한 모형의 실증 검증을 위해 한국 주식시장의 10개 업종 지수 데이터를 활용하여 2003년부터 2019년까지 총 17년간 주가 자료를 활용하였으며 in-sample 1,000개, out-of-sample 20개씩 Moving-window 방식으로 예측 결과값을 누적하여 총 154회의 리밸런싱이 이루어진 백테스팅 결과를 도출하였다. 본 연구에서 제안한 자산배분 모형은 기계학습을 사용하지 않은 기존의 리스크패리티와 비교하였을 때 누적수익률 및 추정 오차에서 모두 개선된 성과를 보여주었다. 총 누적수익률은 45.748%로 리스크패리티 대비 약 5% 높은 결과를 보였고 추정오차 역시 10개 업종 중 9개에서 감소한 결과를 보였다. 실험 결과를 통해 최적화 자산배분 모형의 추정 오차를 감소시킴으로써 포트폴리오 성과를 개선하였다. 포트폴리오의 추정 오차를 줄이기 위해 모수 추정 방법에 관한 다양한 연구 사례들이 존재한다. 본 연구는 추정 오차를 줄이기 위한 새로운 추정방법으로 기계학습을 제시하여 최근 빠른 속도로 발전하는 금융시장에 맞는 진보된 인공지능형 자산배분 모형을 제시한 점에서 의의가 있다.

주제어 : 자산배분 모형, 리스크패리티, 포트폴리오 최적화, XGBoost, 추정 오차

논문접수일 : 2020년 2월 12일 논문수정일 : 2020년 3월 11일 게재확정일 : 2020년 3월 15일
원고유형 : 일반논문 교신저자 : 최흥식

1. 서론

현대사회는 과학기술 및 인터넷의 발전으로 인해 점점 더 고도화되고 있다. 구글, IBM, 마이크로소프트 등 글로벌 기업들은 인공지능 개발에 참여하고 있으며 신약 개발, 의료기술, 자율주행차 등 수 많은 분야에서는 패러다임의 변화(paradigm shift)가 일어나고 있다. 금융산업 역시 인공지능의 도입을 위해 많은 연구가 진행되고 있으며 금융시장에 로보-어드바이저(Robo-advisor)를 도입한 금융상품들이 출시되고 있다. 로보-어드바이저란 Robot과 Advisor의 합성어로 재무설계나 투자 결정을 금융 공학적 알고리즘으로 자동화하여 온라인으로 제공하는 투자자문업자를 지칭한다(Sa et al., 2016). 최근 로보-어드바이저는 자산배분 모형과 결합한 형태로 발전되고 있다.

Markowitz(1952)는 이전까지 추상적으로 존재하던 분산투자의 효율성을 수학적으로 증명하여 현대 포트폴리오 이론을 체계화하였다. 평균분산 모형, 블랙리터만 모형, 리스크패리티 모형 등 다양한 자산배분 모형이 등장하였고 Brinson (1995)은 포트폴리오 성과가 종목 선택, 마켓타이밍 그리고 자산 배분 중 91.5%가 자산 배분을 통해 결정된다는 연구 결과를 발표하였다. 이처럼 오늘날 자산 배분의 중요성은 강조되고 있다.

Markowitz가 처음 제시하며 최적 자산배분 모형의 기반이 된 평균분산모형은 효율적 투자선(efficient frontier) 상에서 위험 대비 기대수익률이 높은 포트폴리오를 선택함으로써 최적의 포트폴리오를 선택하는 모형이다. 평균분산모형은 자산의 과거 수익률을 기반으로 계산된 평균, 표준편차 및 상관계수를 활용해 간단하게 계산할 수 있어 적용이 쉽다는 장점이 있다.

그러나 투자 비중에 대한 최적 해를 구하는 과정에서 타 자산 대비 기대수익률이 높고 위험구조가 다른 자산에 지나치게 많은 투자 비중이 계산되고 음의 기대수익률의 자산에는 투자 비중이 0이 되는 코너 해(corner solution)가 발생하게 된다. 이 점을 보완하기 위해 Black and Litterman (1991)은 시장 포트폴리오로부터 균형기대수익률을 도출하여 코너 해 문제를 해결하고 투자자 전망을 적용한 블랙리터만 모형을 제시하였다. 또한 Qian(2005)은 포트폴리오에 속한 자산들의 위험기여도를 동일하게 맞춰 특정 자산에 의해 위험이 과편중 되지 않도록 설계한 리스크패리티 모형(risk parity model)을 제안하였다. 오늘날 수많은 투자자들은 이와 같은 이론을 바탕으로 포트폴리오 전략을 수립하고 있다.

하지만 자산배분 모형은 최적화 과정에서 발생한 추정 오차로 인해 많은 문제점을 야기한다. Chopra and Ziemba(1993)는 모수 추정치의 작은 변화에도 최적 포트폴리오 구성에 큰 변화를 야기할 수 있다고 발표하였다. 그리고 Demiguel et al.(2009)은 자산배분 모형의 모형 오차나 추정 오차로 인해 자산배분 성과에 상당한 영향을 끼친다고 주장하였다. 국내 연구로는 Lee and Lee (2002)는 모수 추정치의 오차가 평균분산모형에 미치는 영향을 연구하여 기대수익률의 추정 오차가 공분산 추정 오차에 비해 훨씬 심각한 영향을 끼친다고 발표하였다. 추정 오차로 인한 최적 포트폴리오의 단점을 개선하기 위해 Lee(2014)는 역사적 공분산 이외의 다양한 공분산 추정방법을 제시하여 추정 오차를 줄이기 위한 모형들을 제시하였다.

본 연구에서도 공분산의 추정 오차를 줄이기 위한 새로운 방법으로 기계학습을 통한 예측으로 모형의 추정치를 대신하고자 한다. 최근 금융 분

야에도 기계학습을 적용한 다양한 사례들이 존재한다. Roh(2013)는 전통적으로 금융 시계열 데이터 분석에 많이 활용되는 ARCH(AutoRegressive Conditional Heteroscedasticity)와 GARCH(Generalized AutoRegressive Conditional Heteroscedasticity)에 ANN(Artificial Neural Network)을 결합시켜 단일 시계열보다 변동성 방향 예측에 성능을 개선한 인공지능형 금융 시계열 예측 모형을 제시하였다. Kim(2019)은 블랙리터만 모형의 투자자 전망을 애널리스트 시장 전망 대신 SVM(Support vector Machine)을 활용하여 객관화된 지능형 자산배분 모형을 제시하였다. Ru and Shin(2011)은 ANN을 활용하여 KOSPI200 지수옵션의 동적 델타 헷징 전략을 제시하였고 기존의 블랙-숄즈 기반의 동적 델타 헷징 방법론보다 비용적 측면에서 높은 개선 효과를 보였다. Ha et al.(2019)는 XGBoost(eXtreme Gradient Boosting)를 활용해 KOSPI20 주가지수 등락 예측하여 시계열 분석에 강점을 가진 LSTM(Long Short Term Memory)과 AR(Auto Regression)의 예측 결과를 비교하여 XGBoost가 금융 시계열 분석에 있어 유용성이 있음을 보여주었다. 그러나 Ha et al.의 연구 외에는 국내에서 XGBoost와 금융모형을 결합한 사례를 찾아보기 힘들다. 이에 본 연구도 기계학습 방법론 중 XGBoost과 전통적 방식의 자산배분 모형을 결합한 모형을설계하고자 한다.

최적 자산배분 모형 중 리스크패리티 모형은 대표적인 위험 기반 자산배분 모형으로 기대수익률을 추정하지 않고 오직 공분산만 추정한다. 상대적으로 추정 오차로 인한 영향이 적은 공분산만을 추정하기 때문에 리스크패리티는 평균분산모형에 비해 안정성이 높다. 또한 위험 회피적 성향을 갖는 모형으로 자산배분의 기본적인 목적과도 부합한다.

이에 본 연구는 자산배분 모형 중 리스크패리티의 추정오차를 줄여 기존에 리스크패리티가 가지는 장점을 살리고 좀 더 진보된 형태의 자산배분 모형을 제시하고자 한다.

2. 이론적 배경

2.1 자산배분 모형

포트폴리오의 위험에는 두 가지 종류가 존재한다. 하나는 체계적 위험(systematic risk)으로 포트폴리오 구성 자산들이 속한 경제 환경에 전체적으로 영향을 주는 분산 불가능한 위험을 말한다. 다른 하나는 비체계적 위험(unsystematic risk)으로 각 자산이 가지는 위험을 뜻한다. 비체계적 위험은 위험구조가 다른 자산들에 분산 투자하여 상쇄시키거나 축소, 제거가 가능하다.

자산배분 모형은 비체계적 위험을 줄이기 위한 계량 모형으로 투자자의 목적 및 성향에 따라 다양한 자산에 투자하여 투자위험을 줄이고 기대수익률을 높인다. 자산배분 모형은 Markowitz의 현대 포트폴리오 이론을 기반으로 발전하였으며 오늘날 많은 기관 투자자들이 자금 운용 목적으로 자산배분 전략을 수립하고 있다.

2.2 리스크패리티

리스크패리티 모형은 Qian(2005)이 처음 제안하였다. 자산들의 위험을 동일하게 맞추게 되어 특정 자산에 의한 위험 편중을 막을 수 있으며 투자 비중이 0이 되는 자산이 나타나지 않아 기존의 평균분산모형이 가지는 코너 해 문제가 발생하지 않는다는 장점이 있다. 또한, 자산의 위험만을 추정하므로 기대수익률 추정으로 야기될

문제들을 사전에 차단할 수 있어 실무적으로 적용하기 용이하다.

리스크패리티 모형은 포트폴리오의 구성 자산이 포트폴리오 전체 위험에 미치는 영향을 위험기여도(risk contribution)라 정의하여 구성 자산들의 위험기여도를 동일하게 맞추는 최적의 투자 비중을 산출한다. 자산의 비중이 한 단위 변동할 때 포트폴리오의 위험이 변동하는 정도를 한계 위험 기여도(marginal risk contribution)라 하며 아래와 같이 투자 비중에 대한 포트폴리오 위험의 편미분으로 표현할 수 있다.

$$mrc_i = \frac{\partial \sigma_p}{\partial w_i}$$

mrc_i : asset(i)`s marginal risk contribution
σ_p : risk of portfolio
w_i : investment weight of asset(i)

위험기여도는 한계 위험기여도와 자산 비중을 곱하여 다음과 같이 표현할 수 있으며 자산군의 위험기여도의 합은 포트폴리오 위험과 동일하다.

$$rc_i = \frac{\partial \sigma_p}{\partial w_i} \times w_i \quad \text{or} \quad mrc_i \times w_i$$
$$\sum_i^n rc_i = \sum_i^n \left(\frac{\partial \sigma_p}{\partial w_i} \times w_i \right) = \sigma_p$$

rc_i : asset(i)`s risk contribution

리스크패리티의 투자 비중 도출 과정은 공분산 행렬을 통해 산출할 수 있으며 각 자산들의 위험기여도 차이가 최소가 되는 시점의 투자 비중을 산출한다. 해당 과정은 다음과 같은 로직으로 표현할 수 있다.

```
Algorithm 1: Risk Parity asset allocation

Input: covariance_matrix
Output: weight

float sigma, variance, RC_diffs, asset_count
matrix wight, covariance_matrix, mrc, rc

Function RC(weight, covariance_matrix):
    variance = weight.T @ covariance_matrix @ weight
    sigma = √variance
    mrc = 1/sigma X (covariance_matrix @ weight)
    rc = weight X mrc
    rc = rc / sum(rc)
    return rc

Init weight = array 1/len(asset_count)
float tolerance = 1e-20

While RC_diffs < tolerance do
    constraints = [sum(weight) = 1 and weight > 0]
    mat_RC = flattening(RC(weight)-RC(weight).T)
    minimize  RC_diffs = Σ √mat_RC²
```

〈Figure 1〉 Risk parity algorithm

2.3 XGBoost 모형

XGBoost(eXtreme Gradient Boosting)는 Chen and Guestrin(2016)이 처음 제시하였으며 의사결정나무(Decision Trees)의 부스팅(boosting) 기법을 개선한 모형이다. 부스팅이란 수많은 약한 분류기를 결합하여 강한 분류기를 만드는 앙상블(ensemble) 기법의 하나로 약한 트리 분류기를 순차적으로 실행하면서 이전 단계의 오차를 보완하는 방식으로 강한 분류기를 생성한다. 하지만 기존 부스팅 기법은 순차적 학습으로 인해 느리다는 단점을 가진다. XGBoost는 이런 부스팅의 느린 학습속도를 병렬화를 통해 개선하여 데이터 분석 경쟁 플랫폼인 캐글(http://www.kaggle.com)에서도 우수성을 입증하며 유명세를 떨치고 있다.

XGBoost는 트리생성 시 CART(Classification And Regression Trees)모형을 사용한다. CART 알고리즘은 의사결정나무의 가장 대표적인 모형

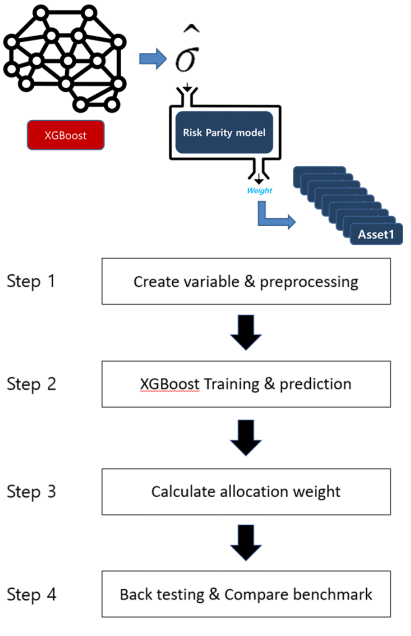
으로 범주형, 연속형 모두 학습이 가능하며 이진 분류를 통해 학습한다. 분류기준으로 불순도 척도인 지니계수(Gini index)를 사용한다. CART 방식은 트리의 모든 리프들이 모델의 최종 스코어에 연관되어 같은 분류 결과를 가지는 트리 모델끼리도 우위를 비교할 수 있다. 생성된 트리 분류기들은 정확도 스코어를 측정해 각 순서에 따라 강한 분류기에서 약한 분류기까지 무작위로 생성하고 이를 순차적으로 개선에 나가며 강력한 트리 분류기를 생성한다.

XGBoost의 가장 큰 장점은 제한된 메모리 환경에서도 수십억 가지의 예제로 확장이 가능할 뿐만 아니라 이를 병렬처리를 통해 학습속도가 매우 빠르다는 점이다. 약한 분류기를 개선해 나가는 과정에서 탐욕적 알고리즘(greedy algorithm)을 사용하여 분류 모델들을 발견하고 분산처리를 통해 최적의 비중을 계산한다. 이 과정에서 트리 분류기의 불필요한 부분을 제거하는 가지치기를 통해 모형의 과적합화를 사전에 방지할 수 있다.

XGBoost 모형은 파라미터 조정을 통해 데이터의 특성과 모형의 용도에 맞는 모형 활용이 가능하다. 파라미터의 종류로는 크게 세분류로 나눌 수 있다. 우선 일반 파라미터로 모형의 전체적인 기능을 설정할 수 있으며 부스터의 종류, 병렬처리에 활용될 스레드 숫자를 선택할 수 있다. 두 번째는 부스팅 파라미터로 모형에서 활용되는 트리 부스터의 설정값을 제어하게 된다. 학습률(learning rate), 트리의 깊이, 트리 당 노드 개수 등 트리의 구조에 관여한다. 마지막으로 학습과정 파라미터는 모형의 최적화 과정을 설정할 수 있다. 대표적으로 목적함수가 존재하며 데이터의 구조나 추출하고자 하는 출력 변수에 따라 다양한 설정이 가능하다.

3. 제안 모형

본 연구에서는 XGBoost로 예측한 표준편차를 활용하여 예측 공분산 행렬을 새롭게 가정하였고 이를 기존의 리스크패리티의 입력 변수인 공분산 행렬 대신 입력하여 모형을 설계하였다. 본 모형은 XGB-RP(XGBoost-Risk Parity)라 명하였으며 과거를 통한 추정이 아닌 투자 기간의 변동성을 예측함으로써 위험을 회피하고 추정오차를 줄여 모형의 안정성을 더한 모형이다. 실증 분석을 위해 국내 업종별 지수(FnGuide Universe 10: 에너지, 소재, 산업재, 경기소비재, 필수소비재, 의료, 금융, IT, 통신서비스, 유틸리티)를 활용하였다. XGBoost 학습 설계, 리스크패리티 투자 비중 산출 및 백테스팅은 파이썬(python)을 통해 구현하였다. 모형의 전체적인 구조는 <Figure 2>와 같이 설계하였다.



〈Figure 2〉 XGB-RP's structure and process

3.1 변수 설정

데이터 가이드(DataGuide5)를 통해 얻은 시가, 고가, 저가, 종가, 거래량, 거래대금, 외국인 보유 비중, 외국인 보유 주식 수를 기초로 하여 <Table 1>과 같은 기술적 지표를 파생하여 입력 변수로 향후 20영업일 간 변동성을 출력 변수로 활용하였다. 기술적 지표 중 변동성, 방향성, 거래량과 관련된 지표들을 입력 변수로 활용하였다.

3.2 XGBoost 파라미터 설정

본 모형의 XGBoost는 파이썬의 'xgboost' 라이브러리를 활용하여 구현하였다. <Table 2>와 같이 파라미터를 설정하였고 그 외 파라미터는 기본값으로 설정하였다.

〈Table 2〉 XGBoost model parameter settings

Category	Parameter	Value
General parameter	booster	gbtree
	silent	None
	nthread	default
Booster parameter	eta(learning_rate)	0.1
	max_depth	2
	subsample	0.8
Learning task parameter	objective	reg:linear

〈Table 1〉 Input and output variables

No.	Variable name		Description
1	Input	Close	Close price
2		Volume	Trading volume
3		Amount	Trading amount
4		Foreign_rate	Foreign ownership ratio
5		Foreign_shares	Number of shares held by foreigners
6		Rate	Rate of return
7		STD	Standard deviation
8		TR	True range
9		NATR	Normalized Average True Range
10		MA5	5day Moving average
11		MA15	15day Moving average
12		MA30	30day Moving average
13		Upper_band	Top line of Bollinger band (time period = 20, std = 2)
14		Middle_band	Middle line of Bollinger band (time period = 20, std = 2)
15		Lower_bnad	Bottom line of Bollinger band (time period = 20, std = 2)
16		Band_width	Interval size of Bollinger band of top and bottom
17		Momentum	Return of 10days
18		OBV	On Balance Volume
19	Output	Target_STD	Standard deviation over the next 20days

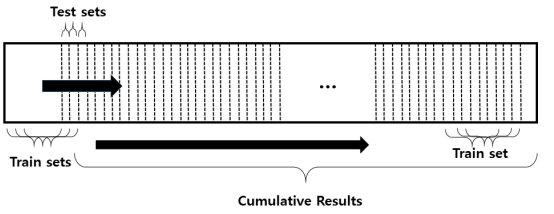
booster의 종류로는 gbtree로 설정하였다. xgboost 모형의 부스터 중에서 가장 기본적인 부스터로 이미 가장성능이 뛰어나다 알려져 있으며 본 모형에서 역시 트리 기반 구조를 활용하기 위해 활용하게 되었다. nthread는 병렬처리에 활용될 스레드 개수를 설정하는 것이며 기본값은 스레드의 최대 수로 활용하는 것이다. 본 모형에서 역시 기본값으로 설정하여 최대 병렬화를 통해 모형의 학습속도를 올렸다.

eta는 기본값 0.3 보다 낮게 잡아 부스팅 프로세스를 좀 더 보수적으로 설계하였고 그 외 과최적화의 예방하고자 max_depth와 subsample을 제어하였다. max_depth는 트리의 깊이가 깊어질수록 subsample은 1과 가까울수록 과최적화에 노출되기 때문에 일정 단위로 수치를 변경해가며 최적화 과정을 실시하여 설정값을 도출하였다.

objective 파라미터는 데이터 구조 및 특성에 맞게 설정할 수 있다. 대표적으로 reg:linear, reg:logistic, binary:logistic, multi:softmax, multi:softprob 등이 있으며 본 모형에서는 이진 분류 혹은 다변량 분류가 아닌 특정 수치를 예측하기 때문 'reg:linear'를 사용하였다.

3.3 Moving-window 방식의 모형 학습

본 모형은 학습의 예측 결과보다는 백테스팅을 통한 결과에 중점을 두었다. 장기간의 백테스팅 결과를 얻기 위해 Moving-window 전진적 분석방법을 적용하여 예측 결과를 누적하였다. <Figure 3>과 같이 1,000영업일 학습, 20영업일 예측을 반복 시행하여 누적된 결과를 도출하였고 실험 데이터 10종 모두 같은 방식으로 진행하였다.



〈Figure 3〉 Structure of Moving window

3.4 예측 공분산 행렬

본 모형은 투자 비중의 산출을 위한 입력 변수로 공분산 행렬이 필요하다. 그러나 기존에 사용하던 자산 간의 상관계수와 표준편차로 구하는 공분산 행렬이 아닌 XGBoost로 도출한 예측 표준편차를 넣은 새로운 방식의 공분산 행렬을 넣기 위해 다음과 같은 방식으로 예측 공분산 행렬을 도출하였다.

$$cov(a,b) = \rho(a,b) \times \hat{\sigma}_a \times \hat{\sigma}_b$$

$cov(a,b)$: covariance between asset A and B
$\rho(a,b)$: correlation between asset A and B
$\widehat{\sigma_a}$: predicted standard deviation of asset A
$\widehat{\sigma_b}$: predicted standard deviation of asset B

위 식의 방식으로 전체 자산에 적용하여 예측 공분산 행렬을 도출한다. 해당 과정은 다음 수식과 같이 표현할 수 있다. 최종적으로 계산된 예측 공분산 행렬은 본 연구에서 제안하는 XGB-RP의 입력 파라미터로 활용된다.[1)]

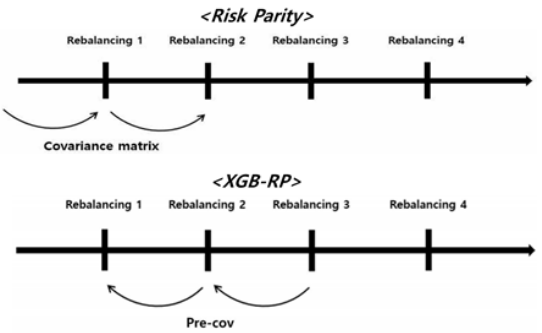
1) 행렬 연산 중 각 성분끼리 곱하는 아다마르 곱(hardamard product)을 활용.

$$\begin{pmatrix} \widehat{\sigma_1} \\ \vdots \\ \widehat{\sigma_n} \end{pmatrix} \times \begin{pmatrix} \widehat{\sigma_1} \cdots \widehat{\sigma_n} \end{pmatrix} = \begin{pmatrix} \widehat{\sigma_1}\widehat{\sigma_1} \cdots \widehat{\sigma_1}\widehat{\sigma_n} \\ \vdots \quad \ddots \quad \vdots \\ \widehat{\sigma_n}\widehat{\sigma_1} \cdots \widehat{\sigma_n}\widehat{\sigma_n} \end{pmatrix}$$

$$\begin{pmatrix} \widehat{\sigma_1}\widehat{\sigma_1} \cdots \widehat{\sigma_1}\widehat{\sigma_n} \\ \vdots \quad \ddots \quad \vdots \\ \widehat{\sigma_n}\widehat{\sigma_1} \cdots \widehat{\sigma_n}\widehat{\sigma_n} \end{pmatrix} \circ \begin{pmatrix} \rho(A_1, A_1) \cdots \rho(A_1, A_n) \\ \vdots \quad \ddots \quad \vdots \\ \rho(A_n, A_1) \cdots \rho(A_n, A_n) \end{pmatrix}$$

$$= \begin{pmatrix} \widehat{\sigma_1}\widehat{\sigma_1}\,\rho(A_1, A_1) \cdots \widehat{\sigma_1}\widehat{\sigma_n}\,\rho(A_1, A_n) \\ \vdots \quad \ddots \quad \vdots \\ \widehat{\sigma_n}\widehat{\sigma_1}\,\rho(A_n, A_1) \cdots \widehat{\sigma_n}\widehat{\sigma_n}\,\rho(A_n, A_n) \end{pmatrix}$$

전통적 방식의 리스크패리티는 과거 기간의 공분산 행렬을 통해 미래 투자 비중을 추정하지만 XGB-RP는 미래 투자 시점을 예측하여 추정 시점과 투자 시점을 동일하게 맞춰 괴리를 줄였다. <Figure 4>는 두 모형의 추정 시점을 도식화하여 이해하기 쉽게 표현하였다.



〈Figure 4〉 Asset allocation diagram

3.5 백테스팅 프로세스

우선 자산배분 비중을 구하기 위해 <Figure 1>의 리스크패리티 투자 비중 산출 과정에 예측 공분산 행렬을 대입한다. 본 과정에서 모든 자산의 위험기여도를 정확하게 일치할 수 없어 위험기여도 차의 합이 일정값 이하로 감소하였을 때의 비중을 최종 투자 비중으로 설정하였다. 최적화 과정에서 초깃값을 동일 비중(1/n)으로 설정하였고 점차 위험기여도 차의 합이 감소하는 방향으로 최소화(minimize) 함수를 통해 최적화하였다.

다음으로 산출된 투자 비중으로 리밸런싱 기간에 맞춰 투자 백테스팅을 진행한다. 전기 말의 자금을 다시 도래한 리밸런싱 시기의 투자 비중에 맞춰 재분배하는 방식으로 투자를 최종 리밸런싱 기간까지 반복 시행하여 최종적으로 누적 수익률을 산출한다. <Figure 5>는 산출된 투자 비중에 맞추어 기말 자금을 재배분하는 과정을 예제로 표현하였다.

example of asset allocation back-testing

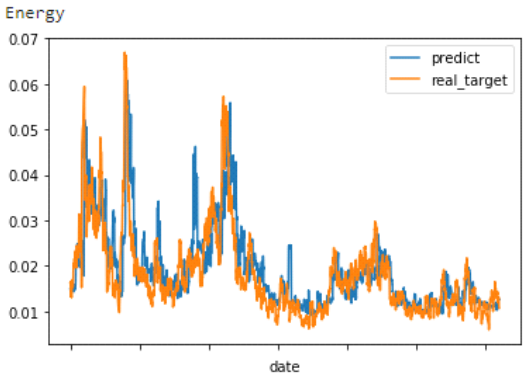
	weight of Investment			Inventory	
Period	Asset A	Asset B	Asset C	Beginning inventory	Ending inventory
1	0.30	0.20	0.50	₩ 100,000.00	₩ 130,000.00
2	0.25	0.35	0.40	₩ 130,000.00	₩ 135,000.00
3	0.15	0.50	0.35	₩ 135,000.00	₩ 153,000.00
4	0.30	0.25	0.45	₩ 153,000.00	₩ 170,000.00

〈Figure 5〉 Example of asset allocation process

다음 기까지의 재투자 기간은 XGBoost의 예측 기간과 동일하게 맞춰 20일로 설정하여 2007.02.23.부터 2019.08.19.까지 총 154회 리밸런싱으로 자산배분 백테스팅을 진행하였다.

4. 실험 결과

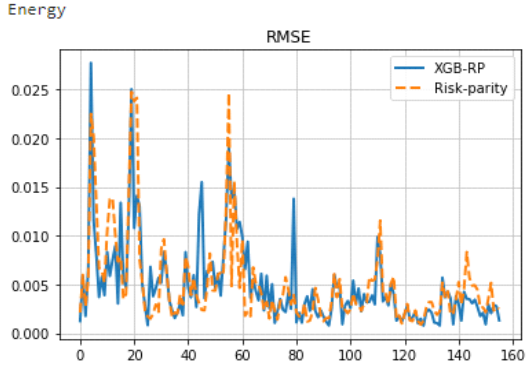
본 모형의 유용성을 검증하기 위해 전통적 리스크패리티와 XGB-RP의 누적수익률 및 추정 오차를 비교하여 성과 분석을 진행하였다. 우선 모형 간의 추정 오차 비교를 위해 XGBoost로 예측한 결과와 실측치 간의 비교를 진행하였으며 <Figure 6>는 10개 업종 중 대표로 에너지 업종의 예측치와 실측치를 비교한 그래프이다. 또한



〈Figure 6〉 Prediction and real volatility(Energy)

에너지 업종 이외 9개 업종에서 역시 비슷한 결과를 나타내었다.

<Figure 7>은 리스크패리티와의 상대적 비교를 위해 평균 제곱근 오차 (Root Mean Square Error, RMSE)를 나타낸 그래프이다. <Figure 7>에서 XGB-RP는 예측치와 실측치 간의 RMSE 값을, 리스크패리티는 과거 추정치와 실측치 간의 RMSE 값을 활용하였고 매 리밸런싱 기간마다 두 자산배분 모형을 비교하였다. 또한, <Figure 7>은 10개 업종 중 대표적으로 에너지 업종의 RMSE 비교 결과를 나타내었다.



〈Figure 7〉 RMSE of XGB-RP and Risk parity(Energy)

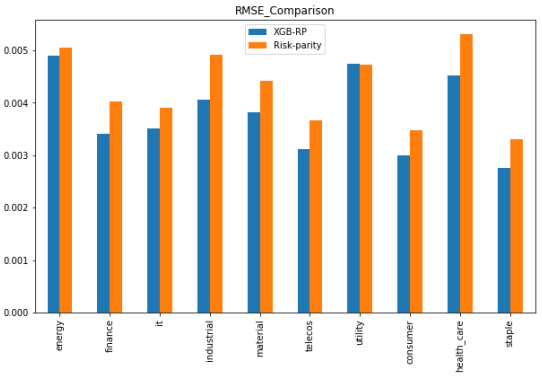
<Figure 7>의 결과 외 타업종에서도 비슷한 결과를 보여주고 있다. 해당 결과를 객관화된 수치로 비교하기 위해 XGB-RP와 리스크패리티의 업종별 전체 기간 RMSE의 평균으로 비교하였다.

<Table 3>는 두 모형의 RMSE 평균을 업종별로 비교한 결과이며 리스크패리티는 업종 평균 0.004278의 추정 오차 그리고 XGB-RP는 0.003785의 추정 오차 결과를 나타냈다. 평균적으로 본 연구에서 제안한 XGB-RP가 전통적 방식의 리스크패리티에 비해 낮은 추정 오차를 나타냄으로써 보다 안정적인 자산배분 모형임을 입증하였다.

〈Table 3〉 RMSE comparison of total sector

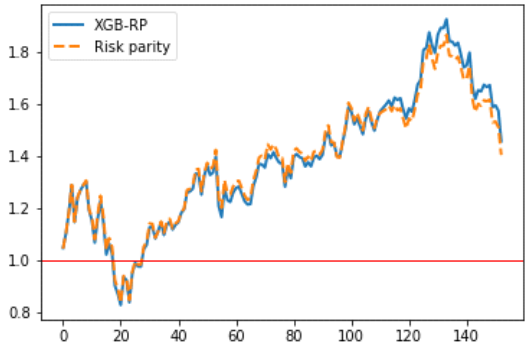
Sector	XGB-RP	Risk parity
Energy	0.004902	0.005043
Finance	0.003405	0.004030
IT	0.003515	0.003895
Industrial	0.004061	0.004911
Material	0.003819	0.004414
Telecos	0.003122	0.003670
Utility	0.004742	0.004730
Consumer	0.002995	0.003471
Health_care	0.004526	0.005314
Staple	0.002765	0.003298
Average	0.003785	0.004278

<Table 3>와 <Figure 8>으로 비교 분석한 결과는 10개 업종 중 유틸리티 업종을 제외한 9개 업종 모두에서 기존의 리스크패리티보다 추정 오차가 감소되었음을 나타내고 있다. 두 결과를 통해 본 연구에서 제시한 모형이 추정 오차를 줄임에 유효성이 있음을 증명하고 있다. 최적화 자산배분 모형에 있어 추정 오차를 줄여 안정적인 모형을 설계하였다.



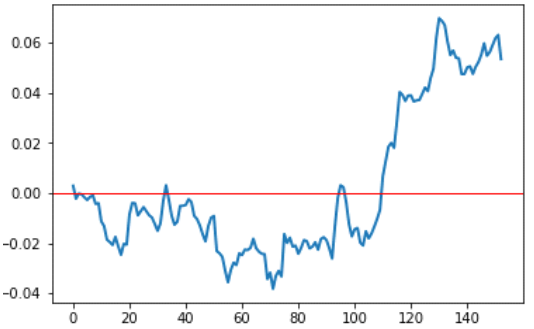
〈Figure 8〉 RMSE comparison of total sector

다음으로 수익률 측면에서의 개선을 확인하기 위해 자산배분 백테스팅을 진행하였다. <Figure 9>은 XGB-RP와 리스크패리티 모형의 누적수익률을 비교한 그래프이다. 제안 모형의 기본적인 원리가 리스크패리티를 기반으로 설계하였기 때문에 두 모형 간의 수익률 측면에서는 큰 차이를 보이지 않지만, 재투자가 반복됨에 따라 누적수익률이 일정 부분 우위를 나타내었다.



〈Figure 9〉 Compounded return of XGB-RP and Risk parity

기간에 따라 두 모형 간 누적수익률의 차이를 시각적으로 확인하기 위해 <Figure 10>을 통해 누적수익률 차이를 나타내었고 시간이 지남에 따라 점차 누적수익률의 차이가 커졌으며 최종적으로 XGB-RP가 리스크패리티 보다 약 5% 수익률 우위를 나타내었다.



〈Figure 10〉 Difference of yield on models

<Table 4>는 두 자산배분 모형이 시뮬레이션 동안 보여준 포트폴리오의 기대수익률, 변동성, 위험조정 수익률 및 누적수익률을 나타낸 표이다. 두 모형의 성과를 비교한 결과는 XGB-RP가 수익률 측면인 기대수익률과 누적수익률에서 기존의 방식보다 개선되었음을 보여준다. 하지만 포트폴리오의 변동성에서는 큰 차이를 보이지 않았다.

〈Table 4〉 Result of asset allocation models

	XGB-RP	Risk parity
$return_p$	45.748%	40.400%
μ/σ	0.079	0.073
μ_p	0.337%	0.312%
σ_p	4.283%	4.278%

$return_p$: cumulative yield of portfolio
μ/σ : risk-adjusted return
μ_p : average of return (expected return)
σ_p : standard deviation of portfolio (risk)

5. 결론 및 연구의 한계점

본 연구에서는 최근 다양한 분야에서 우수성을 증명한 XGBoost와 전통적 자산배분 모형을 결합한 인공지능형 자산배분 모형을 제시하였다. 전통적 자산배분의 한계였던 역사적 데이터로 인한 추정 오차를 기계학습을 통한 예측으로 해결하였다. 또한 추정 기간과 투자 기간의 괴리를 줄여 직관적으로 이해하기 용이하고 추정 오차의 감소로 모형을 안정성을 높였다.

본 모형은 예측된 위험을 미리 회피하도록 설계해 리스크패리티가 가지는 위험 회피적 성향을 강조하였다. 실증 분석을 위한 자료는 2003년부터 2019년까지의 17년 동안의 한국 시장의 10개 업종 지수를 활용해 분석하였다. 예측성능의 한계로 인해 20일 간격의 리밸런싱을 진행하였고 Moving-window 방식의 학습 구조를 설계해 장기간의 out-of-sample 기간을 정용해 총 154회의 재투자한 결과를 도출하였다. 전통적 방식의 리스크패리티와 비교하였을 때 약 5%의 수익률 상승을 보였고 RMSE는 0.004278에서 0.003785로 감소시켰다.

변화하는 금융시장에서 과거 자산의 특징이 미래에도 지속될 것인가에 대한 가장 근본적인 의문으로 인해 자산배분 모형뿐만 아니라 수많은 금융모형은 실무에서 활용되는데 한계를 보이고 있다. 하지만 본 연구는 기존의 자산배분 방식의 장점을 살릴 뿐만 아니라 자산의 위험을 최신 알고리즘으로 예측하고 대응함으로써 전통적 방식의 한계를 보완하고 실무적으로 활용하기 용이하도록 설계하였다. 또한, 추정 오차를 줄여 자산배분 모형의 안정성을 올리는 새로운 공분산 추정방법을 제시하였다. 본 사례를 통해 앞으로 있을 다양한 자산배분 연구에 조금이나마 이론적으로 기여할 수 있다고 생각한다.

하지만 예측 모형의 한계로 인해 먼 미래까지 예측할 수는 없었다. 이에 비례하여 리밸런싱 기간 역시 짧아질 수밖에 없었다. 실제 투자에서 동반될 거래 비용까지 고려하게 된다면 빈번한 리밸런싱으로 인해 비용적 측면에서 문제점을 보일 수 있다. 추후 연구에서 예측 성과를 개선하면서 점차 예측 시점을 늘려간다면 좀 더 현실적으로 적용하기 용이한 모형으로 발전할 것이다.

리스크패리티는 위험구조가 다른 자산에 강점을 가진 모형이다. 그러나 본 연구에서는 한국 시장을 기반으로 실증 분석하여 포트폴리오의 구성 자산들에서 비슷한 위험구조를 보였다. 이후 연구에서는 국내외 다양한 투자 원천들 즉 위험구조가 다른 자산을 포트폴리오 구성으로 활용한다면 자산 배분 측면에서 더 우수한 결과를 볼 수 있을 것이다. 본 연구에서는 XGBoost 모형만 제안하였지만 다른 기계학습방법과의 비교를 통해 더 나은 성과를 보이는 모형에 관한 추가적인 연구가 필요하다.

참고문헌(References)

Black, F. and R. Litterman, “Asset allocation: Combining investor views with market equilibrium”, *Journal of Fixed Income*, Vol1, No.2 (1991), 7~18.

Brinson, G. P., “Determinants of portfolio performance”, *Financial Analysts Journal*, Vol51, No.1 (1995), 133~138.

Chen, T. and C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, *22nd ACM SIGKDD Conference on Knowledge Discovery and*

Data Mining(2016), 785~794.

Chopra, V. K. and W. T. Ziemba, "The effect of errors in means, variances, and covariances on optimal portfolio choice", *Journal of Portfolio Management*, Vol.19, No.2(1993), 6~11.

Demiguel, V., L. Garlappi and R. Uppal, "Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?", *Review of Financial Studies*, Vol.22, No.5(2009), 1915~1953.

Hah, D. W., Y. M., Kim and J. J., Ahn, "A study on KOSPI200 direction forecasting using XGBoost model", *Journal of the Korean Data And Information Science Society*, Vol.30, No.3(2019), 655~669.

Kim, S. W., "Robo-Advisor Algorithm with Intelligent View Model", *Journal of Intelligence and Information Systems*, Vol.25, No.2(2019), 39~55.

Lee, C. S. and K. B., Lee, "A Sensitivity Analysis of the effects of errors in parameter Estimation on portfolio efficiency", *The Korean Journal of Financial Engineering*, Vol.1, No.0(2002), 1~13.

Lee, S. H., "Covariance Estimation and the Effect on the Performance of the Optimal Portfolio", *Journal of the Korean Operations Research and Management Science Society*, Vol.39, No.4(2014), 137~152.

Markowitz, H. M., "Portfolio selection", *Journal of Finance*, Vol.7, No.1(1952), p.77~91.

Qian, E, "Risk parity portfolios: Efficient portfolios through true diversification, *Panagora Asset Management*, September(2005).

Roh, T. H., "Integration Model of Econometric Time Series for Volatility Forecasting", *Korean Management Consulting Review*, Vol.13, No.1(2013), 313~340.

Ru, J. P. and H. J. Shin, "An Option Hedge Strategy Using Machine Learning and Dynamic Delta Hedging", *Journal of The Korea Academia-Industrial cooperation Society*, Vol.12, No.2(2011), 712~717.

Sa, J. H., S. H., Hee and G. Y. Gim, "A study on Utilization of Robo-Advisor in Korea", *Proceedings of Fall Conference on Korea Society of IT Sercvices*(2016), 234~237.

Abstract

A Study on Risk Parity Asset Allocation Model with XGBoost

Younghoon Kim*·HeungSik Choi**·SunWoong Kim***

Artificial intelligences are changing world. Financial market is also not an exception. Robo-Advisor is actively being developed, making up the weakness of traditional asset allocation methods and replacing the parts that are difficult for the traditional methods. It makes automated investment decisions with artificial intelligence algorithms and is used with various asset allocation models such as mean-variance model, Black-Litterman model and risk parity model. Risk parity model is a typical risk-based asset allocation model which is focused on the volatility of assets. It avoids investment risk structurally. So it has stability in the management of large size fund and it has been widely used in financial field. XGBoost model is a parallel tree-boosting method. It is an optimized gradient boosting model designed to be highly efficient and flexible. It not only makes billions of examples in limited memory environments but is also very fast to learn compared to traditional boosting methods. It is frequently used in various fields of data analysis and has a lot of advantages. So in this study, we propose a new asset allocation model that combines risk parity model and XGBoost machine learning model. This model uses XGBoost to predict the risk of assets and applies the predictive risk to the process of covariance estimation. There are estimated errors between the estimation period and the actual investment period because the optimized asset allocation model estimates the proportion of investments based on historical data. these estimated errors adversely affect the optimized portfolio performance. This study aims to improve the stability and portfolio performance of the model by predicting the volatility of the next investment period and reducing estimated errors of optimized asset allocation model. As a result, it narrows the gap between theory and practice and proposes a more advanced asset allocation model.

In this study, we used the Korean stock market price data for a total of 17 years from 2003 to 2019

* Graduate School of Business IT, Kookmin University, Seoul
** Corresponding Author: Heung Sik Choi
Graduate School of Business IT, Kookmin University
77 Jeongneung-ro, Seongbuk-gu, Seoul 136-702, Korea
Tel: +82-2-910-4567, Fax: +82-2-910-4017, E-mail: hschoi@kookmin.ac.kr
*** Graduate School of Business IT, Kookmin University, Seoul

for the empirical test of the suggested model. The data sets are specifically composed of energy, finance, IT, industrial, material, telecommunication, utility, consumer, health care and staple sectors. We accumulated the value of prediction using moving-window method by 1,000 in-sample and 20 out-of-sample, so we produced a total of 154 rebalancing back-testing results. We analyzed portfolio performance in terms of cumulative rate of return and got a lot of sample data because of long period results. Comparing with traditional risk parity model, this experiment recorded improvements in both cumulative yield and reduction of estimated errors. The total cumulative return is 45.748%, about 5% higher than that of risk parity model and also the estimated errors are reduced in 9 out of 10 industry sectors. The reduction of estimated errors increases stability of the model and makes it easy to apply in practical investment. The results of the experiment showed improvement of portfolio performance by reducing the estimated errors of the optimized asset allocation model.

Many financial models and asset allocation models are limited in practical investment because of the most fundamental question of whether the past characteristics of assets will continue into the future in the changing financial market. However, this study not only takes advantage of traditional asset allocation models, but also supplements the limitations of traditional methods and increases stability by predicting the risks of assets with the latest algorithm. There are various studies on parametric estimation methods to reduce the estimated errors in the portfolio optimization. We also suggested a new method to reduce estimated errors in optimized asset allocation model using machine learning. So this study is meaningful in that it proposes an advanced artificial intelligence asset allocation model for the fast-developing financial markets.

Key Words : asset allocation model, risk parity, portfolio optimization, XGBoost, estimated errors

Received : February 12, 2020 Revised : March 11, 2020 Accepted : March 15, 2020
Publication Type : Regular Paper Corresponding Author : Heung Sik Choi

저 자 소 개

김영훈
서울과학기술대학교 안경광학과에서 학사학위를 취득하였고, 국민대학교 비즈니스IT전문대학원에서 트레이딩시스템 전공 석사학위를 취득하였다. 주요 관심분야는 자산배분 전략, 파생상품 설계, 알고리즘 트레이딩, 금융공학, 로보어드바이저 등이다.

최흥식
현재 국민대학교 경영대학 경영정보학부 및 동 대학 비즈니스IT전문대학원 교수로 재직 중이다. KAIST에서 경영과학 석사학위를 취득하였으며 미국 로체스터 대학에서 경영학 석사 및 박사학위를 취득하였다. 관심분야로는 파생상품 시스템트레이딩, 트레이딩계량분석, 옵션 변동성매매, 글로벌 자산배분 등이다.

김선웅
현재 국민대학교 비즈니스IT전문대학원 교수로 재직 중이다. 서울대학교 경영학과에서 경영학사를 취득하고, KAIST 경영과학과에서 투자론을 전공하여 공학석사와 공학박사학위를 취득하였다. 주요 관심분야는 트레이딩시스템, 투자공학, 헤지펀드와 자산운용이다.