

금융 지표와 파라미터 최적화를 통한 로보어드바이저 전략 도출 사례

강민철* · 임규건**

A Case of Establishing Robo-advisor Strategy through Parameter Optimization

Mincheal Kang* · Gyoo Gun Lim**

■ Abstract ■

Facing the 4th Industrial Revolution era, researches on artificial intelligence have become active and attempts have been made to apply machine learning in various fields. In the field of finance, Robo Advisor service, which analyze the market, make investment decisions and allocate assets instead of people, are rapidly expanding. The stock price prediction using the machine learning that has been carried out to date is mainly based on the prediction of the market index such as KOSPI, and utilizes technical data that is fundamental index or price derivative index using financial statement. However, most researches have proceeded without any explicit verification of the prediction rate of the learning data. In this study, we conducted an experiment to determine the degree of market prediction ability of basic indicators, technical indicators, and system risk indicators (AR) used in stock price prediction. First, we set the core parameters for each financial indicator and define the objective function reflecting the return and volatility. Then, an experiment was performed to extract the sample from the distribution of each parameter by the Markov chain Monte Carlo (MCMC) method and to find the optimum value to maximize the objective function. Since Robo Advisor is a commodity that trades financial instruments such as stocks and funds, it can not be utilized only by forecasting the market index. The sample for this experiment is data of 17 years of 1,500 stocks that have been listed in Korea for more than 5 years after listing. As a result of the experiment, it was possible to establish a meaningful trading strategy that exceeds the market return. This study can be utilized as a basis for the development of Robo Advisor products in that it includes a large proportion of listed stocks in Korea, rather than an experiment on a single index, and verifies market predictability of various financial indicators.

Keyword : Robo Advisor, Markov Chain Monte Carlo (MCMC), Optimization, Financial Indicators, Market Forecasts, Artificial Intelligence

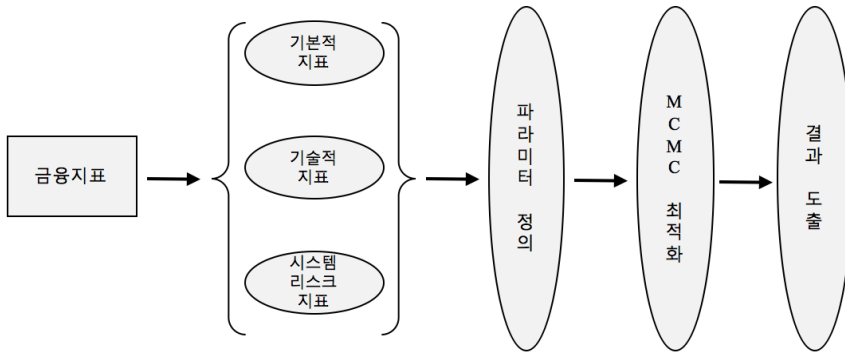
1. 서론

저금리 시대가 도래하자 사람들은 저축 이외에 다양한 금융 상품 투자로 눈을 돌리기 시작하였다. 하지만 전문지식이 없는 일반 투자자들에게 직접 투자나 복잡한 금융 상품은 쉽게 다가가기 힘든 실정이다. 물론 일정 수준 이상의 자산을 보유한 투자자들은 자산 관리사를 고용하여 투자에 대한 자문을 받거나 직접 위탁을 하는 형태로 자산을 관리하는 것이 가능하다. 하지만 시장의 대다수를 차지하는 소액 투자자들에게 자산관리 서비스를 이용하는 것은 부담스러운 일이다. 이러한 투자 여건은 최근 4차산업혁명시대를 맞이하여 로보어드바이저(robo-advisor)에 대한 수요를 증가시키고 있다. 로보어드바이저란 프로그램이 투자자의 개별 정보에 따라 자산의 운용과 배분을 수행하고 수익률을 관리해 주는 서비스를 말한다. 로보어드바이저는 미리 정해진 규칙(rule)으로 자동화된 거래를 수행하는 것이 아니라, 빅데이터를 기반으로 직접 시장을 분석하고 향후 변동되는 시장 상황에 맞추어 포트폴리오를 재구성하는 리밸런싱을 수행한다는 점에서 기존의 자동 거래와 구분된다. 또한 어드바이저(advisor)로서 성립하기 위해서는 투자자 개개인에 맞춤형된 형태로 자산을 운용하는 것이 필수적이다. 이에 대부분의 로보어드바이저 서비스를 제공하는 업체는 투자자의 투자성향을 조사하는 문항을 사전에 실시하여 투자자의 위험성향, 자산 규모, 투자 기간 등을 종합적으로 판단하여 포트폴리오의 위험도를 설정하고 있다. 현재 국내에서는 금융위원회가 2016년 9월부터 시작한 로보어드바이저 테스트베드를 운영하고 있는 상황이다. 참가 기업은 크게 은행, 투자자문사, 그리고 IT 기술 보유 업체로 구분된다.

로보어드바이저의 핵심 분야인 주가 예측은 크게 기본적 분석(fundamental analysis)과 기술적 분석(technical analysis), 그리고 과학기술적 방법(technological methods)으로 구분할 수 있다(김선웅,

안현철, 2010). 기본적 분석은 분기별로 발표되는 기업의 재무제표에 기반하여 기업의 성과를 가늠하여 미래의 주가의 추이를 예측하는 방법이다. 기술적 분석은 개별 주식의 과거 가격 변동을 분석하여 패턴을 찾아내고 그에 기반하여 미래의 가격을 예측하려는 방식이다. 과학기술적 방법은 성능이 크게 진보한 컴퓨터 연산능력과 빅데이터를 분석하여 의미를 추출해내는 데이터마이닝 기법의 발전을 통해 서포트벡터머신(support vector machine)과 인공신경망(artificial neural network)과 같은 기계 학습 알고리즘으로 주식의 변동을 예측하는 방법이다. 하지만 기계학습에 가장 선행되어야 하는 것은 양질의 학습 데이터이다. 실제로 최고의 연산능력과 최신의 알고리즘을 사용하더라도, 학습에 사용되는 데이터가 목표변수에 대한 설명력이 떨어진다면, 기계는 제대로 학습을 하는 것이 불가능하다(Domingos, 2012). 현재까지 진행된 기계 학습을 활용한 주가 예측은 대부분의 앞서 언급한 기본적 지표(fundamental data)와 기술적 지표(technical data)를 활용하였다(박제연, 유재필, 2016).

하지만 이러한 금융 지표들이 주가에 대한 예측력을 지니는가에 대한 실험은 전통적인 머신러닝에서는 훈련집합과 시험집합을 구분하여 일반화된 예측력을 검증하였지만(김동영 등, 2014) 지표 자체에 대한 검증은 부재하였다. 이에 본 연구에서는 이러한 금융 지표들의 주가에 대한 예측력을 확인하고자 한다. 이를 위해서 [그림 1]과 같이 대표적인 기본적 지표인 PBR, PER과 기술적 지표인 볼린저밴드(Bollinger Band)와 MACD(Moving Average Convergence Divergence) 그리고 시장 전체의 리스크를 측정하는 시스템 리스크 지표인 AR(absorption ratio)에 대하여 파라미터를 정의하고, 수익률과 변동성에 대한 목적함수를 정의한다. 그리고, 그것을 극대화하는 파라미터를 마르코프 연쇄 몬테카를로(MCMC)방식으로 최적화하고 최종적으로 수익률에 대한 결과를 도출하여 시장 수익률과 비교하고자 한다.



[그림 1] 금융지표 최적화 실험 절차

2. 선행연구

2.1 금융 지표

주식시장에서 활용되는 주요 금융지표들은 <표 1>과 같이 기본적 지표, 기술적 지표, 시스템 리스크 지표로 나뉘어진다. 기본적 지표는 기업의 재무제표와 가격을 종합하여 도출한 지표이며, 기술적 지표는 주가 데이터를 가공하여 산출되는 변수이다. 마지막으로 시스템 리스크 지표는 시장이 통합되거나 긴밀하게 동조화된 정도를 측정하는 지표이다. 높은 AR은 리스크 요인들이 응집되어 있다는 의미를 나타낸다. 다양한 개별 지표들이 존재하지만 본 연구에서는 대표적으로 사용되는 지표들에

<표 1> 주요 금융지표

구분	주요 지표
기본적 지표 (Fundamental Data)	주가순자산비율(PBR) 주가수익률(PER) 주당순자산가치(BPS) 주당순이익(EPS) 주당EBITDA
기술적 지표 (Technical Data)	변동성 지표 : 볼린저밴드(Bollinger Band) 추세지표 : MACD 모멘텀지표 : 스토캐스틱(Stochastic), RSI(Relative Strength Index) 시장강도 : 거래량이동평균, OBV(On Balance Volume)
시스템 리스크 지표(AR)	주성분분석(PCA)

대해서 실험하고자 한다. 본 장에서는 실험에 사용될 주요 금융지표에 대해서 살펴본다.

기본적 지표에서는 대표적인 주가순자산비율(PBR, Price Book-value Ratio)과 주가수익률(PER, Price Earning Ratio)을 분석한다. PBR은 주가를 주당순자산가치(BPS, Bookvalue Per Share)로 나눈 것이다. 주가가 1주당 순자산의 몇 배로 거래되고 있는가를 표시한다. 주가가 BPS를 초과하면 PBR은 1을 넘어가게 되는데, 이는 시장이 해당 기업의 잠재적 성장률을 주가에 반영하여 거래한다는 의미가 된다. 따라서 PBR이 높을수록, 시장이 해당 기업을 현재 가치 이상으로 고평가한다는 의미를 가진다. PER은 주가를 주당순이익(EPS, Earning Per Share)으로 나눈 것이다. 주가가 1주당 순이익의 몇 배로 거래되고 있는가를 표시한다.

기술적 지표에서는 볼린저밴드(Bollinger Band)와 MACD(Moving Average Convergence and Divergence)를 분석한다. 볼린저밴드는 존 볼린저가 개발한 추세 분석 도구로 이동평균선을 기준으로 주가가 어느정도 위치에 있는지를 측정한다. 중심선, 상단선, 하단선의 3가지 선으로 구성되며, 중심선이 이동평균선이다. 주가가 상단선과 하단선을 돌파하는 것을 신호로 삼는다. 볼린저밴드(John Bollinger, 1980)는 주가의 추세를 파악하면서 매매하는 대표적인 기법이다. 해당 방법은 확률적인 정규분포에 토대를 두고 분석하는 기법이다. 볼린저밴드의 구성은 다음 식과 같다.

$$\begin{aligned} \text{Middle Band} \\ = SMA_t(N) &= \frac{P_t + P_{t-1} + \dots + P_{t-(N-1)}}{N} \end{aligned}$$

$$\begin{aligned} \text{Stochastic Upper Band at } K \text{ times} \\ = SMA_t(N) + K \times STD[SMA_t(N)]_{\text{rolling period } N} \end{aligned}$$

$$\begin{aligned} \text{Stochastic Lower Band at } K \text{ times} \\ = SMA_t(N) - K \times STD[SMA_t(N)]_{\text{rolling period } N} \end{aligned}$$

볼린저밴드는 N기간 동안의 단순이동평균선(또는 지수이동평균선)을 밴드의 중심선으로 하고, N기간 동안의 중심선의 표준편차에 K배(일반적으로 2)한 값을 각각 더하고 빼서 상단선과 하단선을 구한다. 2010년 볼린저밴드의 추가지표가 다음과 같이 소개되었다:

$$\%b_t = \begin{cases} 0 & \text{if } P_t = \text{lower } BB_t \\ 1 & \text{if } P_t = \text{upper } BB_t \\ \frac{P_t - \text{lower } BB_t}{\text{upper } BB_t - \text{lower } BB_t} & \text{else} \end{cases}$$

$$\text{Bandwidth}_t = \frac{\text{upper } BB_t - \text{lower } BB_t}{\text{middle } BB_t}$$

$\%b_t$ 는 t시점에 가격이 밴드의 어느곳에 위치해 있는지 측정할 수 있는 지표이다(Kabasinkas, 2010). 가령 P_t 가 $\text{lower } BB_t$ 와 같다면, $\%b_t$ 는 0을 가진다. 반면, P_t 가 $\text{upper } BB_t$ 와 같다면 $\%b_t$ 는 1을 가진다. Bandwidth_t 는 볼린저밴드의 정규화된 기준으로 밴드의 변동성을 측정하여 추세를 확인하는 데에 이용된다.

MACD는 제라드 아펠이 개발한 지표로 단순이동평균선이 아닌, 최근 주가에 가중치를 부여한 지수이동평균선을 사용한 지표이다(Appel, 2005). MACD란 주가가 변동할 때에 단기 이동평균선과 장기 이동평균선의 차이를 계산한 것이다. 이때 MACD가 그리는 패턴이 수렴(convergence) 혹은 발산(divergence)하는가를 통해서 의미를 해석하는 것이다. 즉, 단기와 장기이동평균이 발산(diverge)하면, 점차 다시 수렴(converge)하려는 속성이 있다는 가정하에 고안된 기술적 지표가 MACD이다.

시스템 리스크 지표는 시장 리스크 요인들의 집중된 정도를 주식시장의 수익률과 공분산 데이터를 주성분분석(PCA, principle component analysis)을 통해 측정하는 것이다(Jung, 2016). 주성분분석은 고차원의 데이터를 저차원의 데이터로 환원시키는 기법이다. 이 경우 주성분분석은 주식 시장과 관련된 여러 가지 변수를 몇 가지 소수의 주성분으로 압축하여 표현하는 것이다. 즉, 이렇게 압축된 주성분들이 시장의 리스크를 많이 설명할수록 시장의 위험요인들이 주성분에 의해 많이 흡수되었다는 의미이므로, 시스템 리스크가 집중되었다고 판단할 근거가 되는 것이다. 이러한 원리로 인해 ‘Absorption ratio(AR)’이라고도 불린다.

2.2 최적화 모형

본 연구에서는 앞서 언급한 금융지표를 변수화하여 수익률과 관련된 목적함수를 설정하고, 각 지표를 이루는 파라미터가 목적함수를 최대화하도록 최적화하는 방식으로 실험을 진행하였다. 이 절에서는 파라미터 최적화에 사용된 마르코프 연쇄 몬테 카를로(Markov Chain Monte Carlo) 방법에 대해 설명하고자 한다.

MCMC 방법은 마르코프 연쇄의 구성에 기반한 확률 분포로부터 원하는 분포의 정적 분포를 갖는 표본을 추출하는 알고리즘의 한 부류이다(Gilks et al., 1996). 큰 수의 단계 이후에 연쇄의 상태는 목표로 하는 분포로부터 추출된 표본처럼 사용될 수 있다. 많은 MCMC 방법은 평형 분포(equilibrium distribution) 주위에서 비교적 작은 보폭으로 움직이며, 이 보폭은 동일한 방향으로 진행하려는 경향을 갖지 않는다. 본 연구는 이와 같은 무작위 행보 MCMC 방법 중 기본적으로 활용되는 메트로폴리스 헤스팅스(MH, Metropolis-Hastings) 알고리즘을 사용하여 목표로 하는 분포에서 표본을 추출하였다. 대부분의 MCMC 알고리즘은 모두 MH 알고리즘의 특별한 경우이거나 MH에서 확장된 알고리즘이다.

메트로폴리스 헤스팅스(MH) 알고리즘은 대상이 되는 각각의 변수의 예측 분포(proposal distribution)와 완전접합밀도함수(full joint density function)를 사용하여 확률 분포에서 표본을 추출한다(Chib et al., 1995). <알고리즘 1>은 MH 알고리즘의 작동 방식을 설명한다.

<알고리즘 1> 메트로폴리스 헤스팅스(MH) 알고리즘

```

Initialize  $x^{(0)} \sim q(x)$ 
for iteration  $i=1, 2, \dots$  do
    Propose :  $x^{cand} \sim q(x^i|x^{i-1})$ 
    Acceptance Probability :
         $\alpha(x^{cand}|x^{i-1}) \min\left\{1, \frac{q(x^{i-1}|x^{cand})\pi(x^{cand})}{q(x^{cand}|x^{i-1})\pi(x^{i-1})}\right\}$ 
     $u \sim \text{Uniform}(u; 0, 1)$ 
    if  $u < \alpha$  then
        Accept the proposal:  $x^i \leftarrow x^{cand}$ 
    else
        Reject the proposal:  $x^i \leftarrow x^{cand}$ 
    end if
end for
    
```

첫 번째 단계는 각 확률 변수를 위한 표본값을 초기화하는 것이다. <알고리즘 1>의 메인 루프는 3가지 부분으로 이루어진다 : (1) 예상 분포인 $q(x^i|x^{i-1})$ 에서 예상 표본인 x^{cand} 를 추출하는 것; (2) 완전접합 밀도함수 $\pi(\cdot)$ 와 예상 분포에 기반하여 허용 함수(acceptance function)를 통해 허용 확률(acceptance probability)을 계산하는 것; (3) 후보 표본을 확률 α 로 수락할지, $1-\alpha$ 의 확률로 거절할지를 결정하는 것이다.

MH 알고리즘은 예상 분포로부터 후보 표본 x 를 추출하고, 허용 확률 α 에 근거하여 후보를 수락할지를 결정한다. 예상 분포(proposal distribution)에는 다음과 같이 대칭과 비대칭의 두 가지 분포가 존재한다.

$$\begin{cases} \text{Symmetric} & \text{if } q(x^i|x^{i-1}) = q(x^{i-1}|x^i) \\ \text{Asymmetric} & \text{if } q(x^i|x^{i-1}) \neq q(x^{i-1}|x^i) \end{cases}$$

대표적인 대칭형 분포로는 마르코프 연쇄의 현재 상태를 중심으로 하는 가우시안 분포(Gaussian distribution)와 균일 분포(uniform distribution)가 있다. 예를 들어, 예상 분포가 가우시안 분포를 가진다고 하면, 후보 표본 x 는 다음과 같다.

$$x^{cand} = x^{i-1} + N(0, \sigma)$$

예상 분포의 대칭성으로 인해 $N(x^{cand} - x^{i-1}; 0, \sigma) = N(x^{i-1} - x^{cand}; 0, \sigma)$ 이기 때문이다. 이러한 예상 분포는 마르코프 연쇄의 현재 상태에 무작위로 노이즈를 발생시켜서 교란된 값을 수락하거나 거절한다(Yildirim, 2012). 이러한 형태의 알고리즘을 무작위 메트로폴리스 알고리즘(Random-walk Metropolis algorithm)이라고 한다.

한편, MH 알고리즘의 허용 함수는 다음과 같은 두 가지 제약 간의 균형을 유지하기 위해 사용된다 : (1) 표본은 완전접합밀도함수 아래에서 더 높은 확률을 지닌 면적에서 추출되어야 한다 : (2) 표본은 공간을 충분히 탐색해야 하며, 특정 공간에 묶여서는 안 된다. 이러한 두 가지 제약이 MH 알고리즘으로 하여금 치우치지 않은 표본을 추출하도록 보장한다. 앞서 언급한 예와 같이 예상 분포가 대칭성을 지니는 경우 허용 함수는 다음과 같이 도출된다.

$$\begin{aligned} \alpha(x^i|x^{i-1}) &= \min\left\{1, \frac{q(x^{i-1}|x^i)\pi(x^i)}{q(x^i|x^{i-1})\pi(x^{i-1})}\right\} \\ &= \min\left\{1, \frac{\pi(x^i)}{\pi(x^{i-1})}\right\} \end{aligned}$$

이처럼 예상 분포가 대칭일 때, 허용 확률은 연쇄의 현재 상태 x^{i-1} 과 예상되는 다음 상태인 x^i 가 얼마나 완전접합밀도함수 아래에 존재하는지에 비례하게 된다. 본 연구에서는 대칭성을 가지는 가우시안 분포의 예상 분포를 사용하여 무작위 메트로폴리스 헤스팅스 알고리즘을 구현하여 각 금융 지표의 파라미터를 최적화하였다.

3. 실험 설계

3.1 자료 선정

본 연구는 KOSPI지수와 같은 단일 지수에 대한 실험이 아닌, KOSPI와 KOSDAQ에 상장된 약 1,500개의 종목 각각에 대한 실험을 진행하였다. 본 연구에서 사용한 실험 데이터는 한국거래소에 상장된 약 1,500개의 일별 주가 데이터와 각 기업의 과거 재무 정보를 연결하는 형태로 가공하였다. 주가 데이터의 시작과 끝은 기본적으로 2000년 1월 1일부터 2017년 1월 6일까지이나 개별 종목 별로 상장 날짜와 거래 정지 등과 같은 이유로 인해 차이가 존재한다.

먼저 주가 데이터는 (주)대신증권의 주가 API를 통해 일괄적으로 수집하였다. 한편 과거 재무 정보 데이터는 기업정보조회 서비스업체인 KISVALUE를 통해 수집하였다. 하지만 주가 데이터는 일별로 구성되는 반면, 기업의 재무 정보는 분기 별로 발표하는 특성으로 인해 분기 별로 구성되어 있었다. 따라서 이 두 주가데이터와 재무데이터를 날짜를 기준으로 병합하는 과정을 거쳤다. 추가로 주식의 거래는 영업일에만 발생하므로, 종합한 데이터의 시계열 색인은 대신증권에서 제공한 영업일을 기준으로 통일하였다.

3.2 기본적 지표 실험 설계

3.2.1 주가순자산비율(PBR)

주가순자산비율(PBR)은 기업의 주가가 해당 기업의 자산가치와 비교해서 어느 정도 비율로 거래되고 있는가를 측정하는 척도이다. 대개 높은 PBR을 지닌 기업은 고평가되어 있다고 평가한다. 본 연구에서는 시계열에서 PBR 지표가 특정 구간 안에 속할 때에 매수하여, 일정 기간이 지난 후에 매도한다고 하였을 때 가장 높은 수익률을 보이는 구간을 추출한다. 해당 기간 중 첫 신호가 발생한 후 20일 이내 동일 신호가 발생할 경우는 매수하지 않았다. PBR에 대한 파라미터는 다음과 같다.

$$Parameter = (c_{from,i}, c_{to,i})$$

여기서 $c_{from,i}$ 는 시계열에서 PBR 구간의 시작점, $c_{to,i}$ 는 구간의 종료점이다. 각각의 파라미터를 다음과 같이 가우시안 분포를 가지는 예상 분포에서 MH 알고리즘을 통해 추출한다. 추출횟수는 300번이다.

$$(x_0 \leq c_{from,i} \leq x_1) \sim N(c_{from,i-1}, 10)$$

$$(y_0 \leq c_{to,i} \leq y_1) \sim N(c_{to,i-1}, 10)$$

초기값은 각각 $c_{from,0} = 0$, $c_{to,0} = 10$ 이다. 그리고 각 파라미터의 범위는 $x_0 = 0$, $x_1 = 5$, $y_0 = 3$, $y_1 = 6$ 이다. 추출된 값을 목적함수에 대입하여 각 보유기간 별로 보유기간 수익률을 구한다. 목적함수는 다음과 같다. 개별 종목 별, 기간 별 테스트해서 목적함수가 최댓값을 갖도록 하는 파라미터를 선택한다. 목적함수는 다음과 같다 :

$$\arg \max_{Parameter} \frac{E[R_{it}]}{STD[R_{it}]} = Sharp \ ratio_t$$

여기서 $E[R_{it}] = w^T R_{it}$ 는 단순 평균이 아닌, 연도를 기준으로 가중한 평균이다. 가중치 w 는 다음과 같이 구한다. 가장 최근 년도는 2017년, 시작 년도는 2000년도로 설정하여 2017년 $w = 18/18 = 1$ 부여. 2016년 $w = 17/18 = 0.9444$, ..., 2000년 $w = 1/18 = 0.055555556$ 을 부여하여 가장 최근 년도에 잘 작동하는 파라미터를 찾는 것을 목적으로 한다(Mamasky, 2000).

표본은 재무데이터가 존재하는 2000년부터 2017년 1월까지 일별 자료를 사용하였으며, 코스피, 코스닥에 상장되어 있는 기업 중 최근일부터 5년 이내에 상장된 기업은 표본에서 제외하였다. 총 표본의 개수는 1,488개 기업이지만, 현재 전략에서 거래 신호가 10번 이상 발생하지 않은 기업은 제거하였다. 분석 시 횡단면 기준 표본의 연 평균 거래 수익률이 연별 표준화한 값의 신뢰구간 5%이하와 95% 이상을 벗어나는 표본은 분석에서 제거하였다.

3.2.2 주가수익률(PER)

주가수익률(PER)은 기업의 주가가 해당 기업의 수익과 비교해서 어느 정도 비율로 거래되고 있는가를

측정하는 척도이다. 따라서 높은 PER을 지니는 기업은 현재 발생하는 수익에 비해 고평가되어 있다고 판단한다. 같은 기본적 지표로서 PBR과 PER은 기업의 정적인 면과 동적인 면을 각각 고려한다는 점에서 상호보완적 측면을 지닌다. 파라미터의 설정과 실험 방법은 3.2.1의 내용과 기본적으로 동일하다.

$$\begin{aligned} \text{Parameter} &= (c_{from,i}, c_{to,i}) \\ (x_0 \leq c_{from,i} \leq x_1) &\sim N(c_{from,i-1}, 10) \\ (y_0 \leq c_{to,i} \leq y_1) &\sim N(c_{to,i-1}, 10) \end{aligned}$$

여기서 초기값은 $c_{from,0} = -1$, $c_{to,0} = 10$ 이다. 그리고 각 파라미터의 범위는 $x_0 = -5$, $x_1 = 3$, $y_0 = 5$, $y_1 = 10$ 이다. 마찬가지로 MH 알고리즘으로 추출된 파라미터값을 입력하여, 목적함수를 최대화하는 값을 찾아낸다. 표본추출횟수는 300번으로 동일하다.

3.3 기술적 지표 실험 설계

3.3.1 볼린저밴드(Bollinger Band)

볼린저밴드는 주가가 밴드 안에서 확장, 수축의 과정을 반복한다고 판단하여 밴드의 변동성이 증가하는 구간은 수축에서 확장으로 전환하는 국면으로 본다. 이때 주가의 움직임이 중심선을 돌파하거나 중심선에서 지지력을 받을 경우 상승추세의 전환으로 판단하여 매수 신호로 판단한다. 본 연구에서는 %b_t가 저점 구간에서 확장할 때 매수하도록 설정하였다. 구체적인 전략은 다음과 같다.

$$\begin{aligned} 0 \leq \%b_{t-3} \leq q_{3,i} \text{ or } 0 \leq \%b_{t-2} \leq q_{3,i} \\ \text{or } 0 \leq \%b_{t-1} \leq q_{3,i} \text{ or } 0 \leq \%b_t \leq q_{3,i} \\ \text{Bandwidth}_{t-2} < \text{Bandwidth}_{t-1} \ \& \ \text{Bandwidth}_{t-3} \\ < \text{Bandwidth}_{t-1} \\ \text{MiddleBand}_{t-1} \times (1 + 0.05) \leq \text{Price}_t \ \& \ \text{Price}_{t-1} \leq \text{Price}_t \end{aligned}$$

이러한 조건을 만족할 경우 t시점의 증가에서 매수하여 정해진 기간 동안 보유하고 매도하는 전략을 정하였다. 볼린저밴드의 파라미터는 다음과 같이 설정하였다.

$$\text{Parameter} = (q_{1,i}, q_{2,i}, q_{3,i})$$

여기서 $q_{1,i}$ 은 단기이동평균선의 기간, $q_{2,i}$ 은 볼린저밴드 표준편차에 곱하여 밴드의 범위를 정하는 K값, 마지막으로 $q_{3,i}$ 은 %b의 상단선이다. 기본적 지표와 마찬가지로 MH 알고리즘으로 각 파라미터의 값의 표본을 추출하여 목적함수를 최대화하는 값을 도출하였다.

$$\begin{aligned} (x_0 \leq q_{1,i} \leq x_1) &\sim N(q_{1,i-1}, 10) \\ (x_2 \leq q_{2,i} \leq x_3) &\sim N(q_{2,i-1}, 1) \\ (x_4 \leq q_{3,i} \leq x_5) &\sim N(q_{3,i-1}, 0.05) \end{aligned}$$

초기값은 $q_{1,0} = 5$, $q_{2,0} = 2$, $q_{3,0} = 0.1$ 이며, 각 파라미터의 범위는 $x_0 = 5$, $x_1 = 60$, $x_2 = 1$, $x_3 = 10$, $x_4 = 0.1$, $x_5 = 1$ 이다. 표본추출은 300번 수행하였다. 목적함수는 기본적 지표에서와 동일하게 sharp ratio로 설정하였다.

3.3.2 MACD(Moving Average Convergence Divergence)

MACD는 최근 주가에 가중치를 부여한 지수이동평균선을 사용하며, 단기지수이동평균선과 장기지수이동평균선의 차이를 지표화한 것이다. MACD는 볼린저밴드와 함께 대표적인 기술적 지표 중의 하나이며, 그 패턴이 수렴 혹은 발산하는 정도를 통해서 매수 혹은 매도의 신호로 사용한다. 설정한 파라미터는 다음과 같다.

$$\text{Parameter} = (q_{1,i}, q_{2,i}, q_{3,i})$$

여기서 $q_{1,i}$ 는 단기지수이동평균선, $q_{2,i}$ 는 장기지수이동평균선, 마지막으로 $q_{3,i}$ 는 매수신호의 이동평균선 n기간이다. 다음과 같은 예상 분포(proposal distribution)에서 마찬가지로 메트로폴리스 헤이스팅스 알고리즘으로 표본을 추출하였다.

$$\begin{aligned} (x_0 \leq q_{1,i} \leq x_1) &\sim N(q_{1,i-1}, 20) \\ (x_2 \leq q_{2,i} \leq x_3) &\sim N(q_{2,i-1}, 500) \\ (x_4 \leq q_{3,i} \leq x_5) &\sim N(q_{3,i-1}, 20) \end{aligned}$$

여기서 초기값은 $q_{1,0} = 12$, $q_{2,0} = 26$, $q_{3,0} = 9$ 이다. 각 파라미터의 범위는 $x_0 = 5$, $x_1 = 20$, $x_2 = 25$, $x_3 = 240$, $x_4 = 1$, $x_5 = 60$ 이다. 목적함수 및 사용된 표본 등의 기타 실험 환경은 불린저밴드와 동일하다.

매수 신호는 t-1시점에 MACD가 전 저점 구간 ± 0.18 구간 이내에 있고, t시점에 MACD가 -1시점보다 상승하면서 t시점에 MACD가 전 저점 이상일 경우 t시점에 가격으로 주가를 보유하여 20일/60일/120일/240일 동안 보유 후 매도하는 전략을 수행하였다. 해당 신호는 t시점 기준 전 5거래일 안에 한번만 발생한 경우에만 매수하는 것으로 하였다.

3.4 시스템 리스크 지표(AR) 실험 설계

시스템 리스크 지표는 주식시장의 주식 섹터가 얼마나 동조화 되어 움직이는지를 나타내는 지표이다. AR의 구성은 아래와 같다.

$$Z_i = \phi_{1,i}x_{1,i} + \phi_{2,i}x_{2,i} + \dots + \phi_{p,1}x_{p,i}$$

여기서 Z_i 는 주식 시장에 영향을 미치는 각 주식 섹터를 i시점을 기준으로 선형결합으로 나타낸 것이다. 주식 섹터는 KSIC 산업 섹터 구분을 기반으로 구성하였다. 이때 섹터 데이터를 구성함에 있어 섹터의 총 수가 62개로 너무 세분화 되었을 뿐 아니라, 섹터에 해당되는 주식이 적어 과거 데이터가 확보가 되지 않는 섹터가 있기 때문에 데이터에 가공이 필요하였다.

방법 1) 과거 관측치가 부족한 12개 섹터만 제거

방법 2) 섹터에 해당되는 주식 수가 25개 이하인 섹터 제거

방법 1)을 사용하였을 경우 가공되는 데이터 길이는 3,609개로 데이터 손실량은 최소가 되지만 남아있는 섹터의 수가 50개가 되어 필요 이상으로 세분화 되어있다. 이 경우 PCA를 통해 주성분을 분석했을 때 주성분의 설명력이 너무 높아지게 된

다. 반면 방법2)를 사용할 경우 가공되는 데이터 길이는 3,546개로 충분한 데이터 크기를 제공하고, 섹터의 수가 20개가 되기 때문에 주성분분석에 활용하기 더 적합하다. 따라서 본 연구에서는 방법2)를 사용하여 섹터 데이터를 가공하였다.

이후에 분류된 주식 당 수익률을 산출하였고, 이를 바탕으로 섹터별 평균 수익률을 산출하였다. 수익률은 연속복리(continuous compound)로 계산하였다. 다음 식은 Z의 분산을 최소화하도록 섹터의 일일 수익률 간의 주성분분석(PCA)을 진행한 것이다. 여기서 K는 주성분의 갯수이다. 이를 바탕으로 AR(K)를 구하고, 최종적으로 표준화를 진행한다.

$$\max\{Var(Z_i)\} = \max_{\phi_{1,i}, \dots, \phi_{p,i}} \left\{ \frac{1}{n} \sum_{k=1}^n \left(\sum_{j=1}^p \phi_{p,i} x_{k,j} \right)^2 \right\}$$

$$AR(K) = \frac{\sum_{i=1}^k Var(Z_i)}{\sum_{j=1}^n Var(Z_j)}$$

$$\Delta AR(K) = \frac{\overline{AR}_{q1} - \overline{AR}_{q2}}{\sigma_{AR,q}}$$

이때, 표준화 된 값인 $\Delta AR(K)$ 이 특정 시그마 수준을 넘었을 경우 AR이 기준보다 크게 증가하였음을 인지한다. 이 경우, 시장이 평소보다 더 동조화 되어있음을 의미하고, 이에 따라 시장의 변동성이 커지게 된다. 따라서 $\Delta AR(K)$ 을 사용한 지표가 시장의 움직임보다 선행되어 변동성의 증가를 예측할 수 있다면, 이를 이용하여 자산배분전략을 효과적으로 운용할 수 있다. 해당 시그널의 발생 판단 기준은 다음의 식을 사용한다.

$$\Delta AR(K) = \frac{\overline{AR}_{q1} - \overline{AR}_{q2}}{\sigma_{AR,q}} > \text{Sigma}$$

위 식을 계산하는 과정에서 필요한 파라미터는 <표 2>와 같다. 기본적 분석과 기술적 분석에 비하여 상대적으로 많은 파라미터를 사용하여서 표로 정리하여 제시한다.

<표 2> 시스템 리스크 지표의 파라미터 설정

파라미터	정의
K_rolling	PCA 매트릭스를 만드는 과정에 필요한 단위 롤링 기간
K	AR지수 측정의 기준이 되는 K개의 주성분
q1	$\Delta AR(K)$ 계산과정에서 사용되는 최근 AR 동향을 판단하는 기간
q2	$\Delta AR(K)$ 계산과정에서 사용되는 기준 AR 동향을 판단하는 기간
Sigma	$\Delta AR(K)$ 가 기준보다 의미 있게 높은 지를 판단하는 수준

각 파라미터의 초기값은 각각 $K_{rolling,0} = 480$, $K_0 = 2$, $q_{1,0} = 15$, $q_{2,0} = 240$, $Sigma_0 = 1.5$ 이며, 세부적인 설정은 다음과 같다.

$$\begin{aligned}
 (60 \leq K_{rolling,i} \leq 720) &\sim N(K_{rolling,i-1}, 20) \\
 (1 \leq K_i \leq 5) &\sim N(K_i, 5) \\
 (1 \leq q_{1,i} \leq 60) &\sim N(q_{1,i}, 10) \\
 (61 \leq q_{2,i} \leq 720) &\sim N(q_{2,i}, 20) \\
 (1 \leq Sigma_i \leq 2) &\sim N(Sigma_i, 1)
 \end{aligned}$$

Sigma를 제외한 파라미터 값은 정수단위로 추출되고, Sigma는 소수 둘째자리에서 추출된다. 최적화 하는 방식은 동일한 알고리즘으로 각 파라미터의 표본을 추출하여 목적함수를 최대화 시키는 방향으로 진행하였다.

목적함수는 이전과 같이 샤프지수로 한 것이 아니라, 매수 신호가 발생한 시점 간 코스피 변동성(표준편차)의 총 합으로 계산하였다. 매수 신호가 이틀이상 연속으로 발생한 기간을 각각 $a_1, a_2, a_3, \dots, a_k$, 그 기간 동안의 코스피의 표준편차를 각각 $\sigma_1(a_1), \sigma_2(a_2), \dots, \sigma_k(a_k)$ 이라고 할 때, 목적함수는 다음과 같다 :

$$\arg \max_{Parameter} \sum_{i=1}^k \sigma_i(a_i)$$

기간 a_k 가 <표 2>의 파라미터로 구성된 함수이므로, 위 목적함수를 최대화하는 방향으로 파라미터 최적화는 진행한다.

목적함수가 다른 것에서 알 수 있듯이, 시스템 리스크 지표는 앞서 진행한 두 가지 실험인 기본적 지표와 기술적 지표에 대한 실험과 차별된다. AR은 K개의 주성분이 시스템 리스크를 설명하는 정도, 다른 말로는 분산을 포함하는 정도를 나타내고, 그것을 바탕으로 위험 요인의 집중정도를 파악하는 것을 목적으로 한다. 따라서 국내 주식시장을 대표하는 지수인 2000년부터 2017년 1월까지의 일별 코스피 증가 자료를 표본으로 사용하였다.

시스템 리스크 지표인 AR에 대한 실험은 개별 종목에 시행하는 것이 아니라, 코스피지수에 적용하여 시장 전체의 위험 집약도를 파악하여 투자의 보조수단으로 활용하기 위함이다. 이에 이동평균(moving average)을 활용하여 AR시그널이 발생했을 경우 동기화된 시장의 움직임이 상승으로 이어질지 하락으로 이어질지를 분석하는 전략을 수립하였다. 여기서는 한 달(20일)단위의 이동평균(ma20)과 일주일 단위의 이동평균(ma5)을 이용하여 분석하였다. <표 3>은 시스템 리스크 지표의 활용 전략을 표기한 것이다. UP 상승, DOWN은 하락, ENLARGE는 확대, DECREASE는 축소 신호이다.

<표 3> 시스템 리스크 지표의 활용 전략

전략	정의
UP(ma_up)	$\Delta AR(K)$ 시그널+ma5-ma20 > 0일 때
DOWN(ma_down)	$\Delta AR(K)$ 시그널+ma5-ma20 < 0일 때
ENLARGE(maD)	$\Delta AR(K)$ 시그널+ma5 _{t-2} -ma20 _{t-2} < ma5 _{t-1} -ma20 _{t-1} +ma5 _{t-1} -ma20 _{t-1} 가 최근 5일동안 -1.2시그마 보다 클 경우
DECREASE(maC)	$\Delta AR(K)$ 시그널+ma5 _{t-2} -ma20 _{t-2} > ma5 _{t-1} -ma20 _{t-1} +ma5 _{t-1} -ma20 _{t-1} 가 최근 5일동안 1.2시그마 보다 작을 경우

4. 실험결과 및 분석

4.1 기본적 지표 실험 결과

4.1.1 주가순자산비율(PBR) 실험 결과

실험 결과 PBR의 목적함수를 최대화하는 보유기간은 240일이었다. <표 4>는 4가지 보유기간 중에서 240일 동안 보유했을 때의 결과를 나타낸 것이다. 나머지 보유기간(20/60/120일)에 대한 결과는 생략한다. <표 4> 결과표는 개별종목의 결과를 평균한 것이다. 왼쪽의 두 개의 열은 처음에 설정하였던 PBR의 상단과 하단의 최적화된 결과값이다. 즉 PBR의 수치가 (0.223, 1.650) 사이에 있을 때 주식의 가격이 상승할 가능성이 높다는 것이다. Stock return은 최적파라미터를 대입하여 구한 평균 거래 수익률(시간가중)이며, Market return은 동일한 거래기간의 평균 시장수익률(코스피, 시간가중)이다. Optimal object stock은 최적화된 목적함수(Sharp ratio)의 최대값이다. Trading frequency는 연평균 거래 발생횟수이다. PBR의 전략의 경우 동기간 시장수익률과 비교했을 때 5배 이상의 수익률을 나타내며 높은 성과를 보였다.

<표 4> PBR 보유기간 240일 결과표

q1	q2	Stock return	Market return	Optimal object stock	Trading frequency
0.223	1.650	2.763%	0.526%	0.710%	7.354

<표 5>는 PBR 실험 결과에 아래의 CAPM 모델 테스트 적용 결과이다.

$$CAPM \text{ 모델} : E[r_t] - r_{ft} = \alpha + \beta E[(r_t - r_{ft})] + e_t$$

<표 5>의 수치들은 개별종목의 연평균 거래수익률을 y, 동일 거래기간의 연 시장수익률을 x로 하여, 자산가격결정모델(CAPM)을 기반으로 회귀 분석하여 도출되었다. 여기서 베타(Beta)는 개별주식의 수익률이 시장수익률의 변동에 얼마만큼 민감한 반응을 보이는가를 측정하는 척도이다. 한편 알파(Alpha)는 적정 기대수익률과 실제 예상된 수익률의 차이를 구한 것으로 초과수익률이다. Panel

A는 각 년도의 개별종목의 거래수익률을 평균하여 동일거래기간의 시장수익률(KOSPI)과 CAPM테스트를 한 결과표이다. 무위험수익률은 1년 국고채 수익률을 사용하였다. 괄호 안의 값은 t-statistic으로 오차정도를 나타낸다. RSE는 잔차표준오차(Residual standard error)이다. Panel B는 Panel A의 결과의 차이를 분석한 것이다.

<표 5>의 CAPM 테스트 결과, 240일 보유전략이 Alpha 0.974배, Beta 1.478배로 다른 보유기간에 비해 0.659로 가장 높은 시장위험조정수익률(Alpha/Beta)을 보였다.

추가로 t시점에서 전체 거래종목수 중에서 상승한 종목수의 비중 즉, 신호가 발생하였을 때 상승할 확률을 계산하였다.

$$up \ hit \ ratio = \frac{\text{상승종목수}_t}{\text{전체거래종목수}_t}$$

분석결과 PBR 전략의 상승비율은 61.6~63.7% 정도로 일반적으로 종목을 매수하였을 때 상승확률이 40%대인 것을 감안하였을 때, 상대적으로 높은 상승확률을 보였다.

<표 5> PBR 실험 결과에 CAPM 테스트 적용 결과

Panel : A				
Holding period	20days	60days	120days	240days
Alpha	0.043 (5.852)	0.171 (3.782)	0.469 (3.675)	0.974 (3.497)
Beta	1.383 (8.540)	2.331 (4.663)	3.923 (6.259)	1.478 (3.352)
RSE	0.023	0.186	0.511	0.901
Adjusted R^2	0.818	0.565	0.705	0.406
F-Statistic	72.94	21.75	39.17	11.23
Panel : B				
Holding period	20days (t1)	60days (t2)	120days (t3)	240days (t4)
Alpha/t	0.22%	0.29%	0.39%	0.41%
ΔAlpha	0.043	0.128	0.298	0.504
Δt	20	40	60	120
ΔAlpha/Δt	0.22%	0.32%	0.50%	0.42%
Beta/t	0.069	0.039	0.033	0.006
ΔBeta	1.383	0.948	1.592	-2.445
ΔBeta/Δt	0.069	0.024	0.027	-0.020
Alpha/Beta	0.031	0.073	0.120	0.659
ΔAlpha/ΔBeta	0.031	0.135	0.187	-0.206

4.1.2 주가수익률(PER) 실험결과

실험 결과 PER의 목적함수를 최대화하는 보유기간은 120일이었다. <표 6>은 4가지 보유기간 중에서 240일 동안 보유했을 때의 결과를 나타낸 것이다. 왼쪽의 두 개의 열은 처음에 설정하였던 PER의 상단과 하단의 최적화된 결과값이다. PER의 수치가 (1.012, 7.115) 사이에 있을 때 주식 가격이 상승할 가능성이 높다는 결과가 도출되었다. 이는 PBR과 비교하였을 때, 상대적으로 더 높은 수치이며, 그 범위도 더 넓은 것으로 나타난다. PER 최적화 전략을 사용하였을 때 시장수익률보다 약 3배의 수익률을 올릴 수 있는 것으로 나타났다.

<표 7>의 CAPM 테스트 결과, 120일 보유전략이 Alpha 0.304배, Beta 1.601배로 다른 보유기간에 비해 Alpha/Beta가 0.190으로 가장 높은 시장위험

<표 6> PER 보유기간 120일 결과표

q1	q2	Stock return	Market return	Optimal object stock	Trading frequency
1.012	7.115	1.343%	0.409%	0.638%	5.843

<표 7> PER 실험 결과에 CAPM 테스트 적용 결과

Panel : A				
Holding period	20days	60days	120days	240days
Alpha	0.036 (3.093)	0.118 (5.527)	0.304 (5.058)	0.974** (3.497)
Beta	0.793 (2.745)	1.217 (3.988)	1.601 (3.134)	1.478** (3.352)
Adjusted R ²	0.290	0.482	0.355	0.602
F-Statistic	7.536	15.910	9.825	25.490
Panel : B				
Holding period	20days (t1)	60days (t2)	120days (t3)	240days (t4)
Alpha/t	0.18%	0.20%	0.25%	0.18%
ΔAlpha	0.036	0.082	0.186	0.133
Δt	20	40	60	120
ΔAlpha/Δt	0.18%	0.21%	0.31%	0.11%
ΔBeta	0.793	0.424	0.384	1.176
ΔBeta/Δt	0.03965	0.0106	0.0064	0.0098
Alpha/Beta	0.045	0.097	0.190	0.157
ΔAlpha/ΔBeta	0.045	0.193	0.484	0.113

조정수익률을 보였다. PER 전략의 상승확률(up hit ratio)는 20일, 60일, 120일, 240일 각각 67.1%, 54.6%, 63.9%, 63.6%를 보이면서 PBR보다 다소 높은 경향을 보였다.

4.2 기술적 지표 실험 결과

4.2.1 볼린저밴드 실험결과

실험 결과 볼린저밴드 전략의 목적함수를 최대화하는 보유기간은 240일이었다. <표 8>은 4가지 보유기간 중에서 240일 동안 보유했을 때의 결과를 나타낸 것이다. q1은 단기이동평균선의 기간, q2는 볼린저밴드의 표준편차에 K배, q3는 %b의 상단선이다. 볼린저밴드 최적화 전략을 사용하였을 때, 시장수익률보다 약 4배의 수익률을 올릴 수 있는 것으로 나타났다. PBR전략의 결과와 비교했을 때, 수익률은 유사한 것에 비해, 목적함수의 최적화값이 더 낮은 것으로 보아, 볼린저밴드의 변동성이 PBR 전략보다 더 변동성이 높을 것을 알 수 있다. 거래빈도는 연평균 3.6회로 기본적 분석을 통한 전략에 비해 다소 드문 것으로 나타났다.

<표 8> PER 보유기간 120일 결과표

q1	q2	q3	Stock return	Market return	Optimal object stock	Trading frequency
11.021	3.121	0.252	2.303%	0.527%	0.596%	3.612

<표 9>의 CAPM 테스트 결과, 240일 보유전략이 Alpha 0.419배, Beta 1.551배로 Alpha/Beta가 0.270으로 가장 높은 시장위험조정수익률을 보였다. 이는 다른 보유기간의 결과와 비교했을 때 시장의 변화에 대한 취약성을 나타내는 베타지수는 낮으면서도 초과수익률인 알파는 높은 것으로 240일이 다른 전략에 비해 우수함을 나타낸다. 볼린저밴드 전략의 상승확률(up hit ratio)는 20일, 60일, 120일, 240일 각각 71%, 66.7%, 63.5%, 63.2%를 보이면서 전반적으로 기본적 분석을 활용한 전략보다 높은 상승확률을 나타내었다.

〈표 9〉 볼린저밴드 실험 결과에 CAPM 테스트 적용 결과

Panel : A				
Holding period	20days	60days	120days	240days
Alpha	0.043 (7.873)	0.106 (5.124)	0.212 (3.531)	0.419 (3.119)
Beta	0.978 (8.085)	1.298 (5.252)	1.693 (4.042)	1.551 (3.120)
Adjusted R ²	0.801	0.624	0.489	0.368
F-Statistic	65.36	27.58	16.34	9.732
Panel : B				
Holding period	20days (t1)	60days (t2)	120days (t3)	240days (t4)
Alpha/t	0.21%	0.18%	0.18%	0.17%
Δ Alpha	0.043	0.063	0.107	0.207
Δ t	20	40	60	120
Δ Alpha/Δ t	0.21%	0.16%	0.18%	0.17%
Δ Beta	0.978	0.320	0.394	-0.142
Δ Beta/Δ t	0.049	0.008	0.007	-0.001
Alpha/Beta	0.044	0.081	0.125	0.270
Δ Alpha/Δ Beta	0.044	0.195	0.271	-1.457

4.2.2 MACD 실험결과

실험 결과 MACD 전략의 목적함수를 최대화하는 보유기간은 240일이었으나, 수익률은 늘어남에도 시장 민감도를 나타내는 베타가 지나치게 높아지는 경향을 보였다. 따라서 적절한 수준의 베타와 알파를 보유한 60일 보유전략이 가장 우수한 것으로 평가하였다. <표 10>의 q1은 단기지수이동평균선, q2는 장기지수이동평균선, q3는 매수신호의 이동평균선 n의 기간이다. 약 1,500 종목의 주식으로 최적화를 한 결과 단기는 약 10일, 장기는 약 48일을 이동평균선으로 MACD를 구성하는 것이 효과적이라는 것을 알 수 있다. 이는 기존의 보편적인 MACD의 단기, 장기이동평균선의 수치와는 차별되는 결과이다. 연평균 거래횟수는 4.7회로 볼린저밴드 전략보다 거래가 더욱 빈번하게 발생하는 것을 확인할 수 있다. 시장수익률에 비해 이 전략의 수익률이 5배 정도 높게 나타남에도 불구하고, 절대적인 수익률이 낮아 매매전략으로서의 효과는 제한적이라고 판단된다.

〈표 10〉 MACD 전략 보유기간 60일 결과표

q1	q2	q3	Stock return	Market return	Optimal object stock	Trading frequency
9.868	48.626	8.964	0.593%	0.114%	0.336%	4.781

〈표 11〉 MACD 실험 결과에 CAPM 테스트 적용 결과

Panel : A				
Holding period	20days	60days	120days	240days
Alpha	0.02 (3.207)	0.123 (4.198)	0.345 (4.573)	0.452 (3.741)
Beta	0.681 (4.212)	1.083 (3.214)	2.073 (3.718)	2.705 (4.625)
Adjusted R ²	0.5113	0.368	0.445	0.576
F-Statistic	17.74	10.33	13.83	21.39
Panel : B				
Holding period	20days (t1)	60days (t2)	120days (t3)	240days (t4)
Alpha/t	0.10%	0.20%	0.28%	0.18%
Δ Alpha	0.02	0.10	0.22	0.10
Δ t	20	40	60	120
Δ Alpha/Δ t	0.10%	0.25%	0.37%	0.08%
Δ Beta	0.681	0.402	0.990	0.632
Δ Beta/Δ t	0.034	0.010	0.017	0.005
Alpha/Beta	0.029	0.114	0.166	0.167
Δ Alpha/Δ Beta	0.029	0.256	0.224	0.169

<표 11>에 의하면, 단순한 Alpha/Beta가 수치는 240일 보유전략이 0.167로 근소하게 높지만, 여기서 120과 240일의 베타값이 2를 넘어서는 것을 볼 수 있다. 이는 해당 전략을 사용했을 경우 시장의 변동성에 2배 이상의 변동성이 발생한다는 것을 나타낸다. 따라서 비록 초과수익률인 알파는 0.123으로 중간 정도이지만, 베타의 수치가 1.083으로 비교적 안정적인 수치를 보이는 60일 전략이 MACD 전략 중에서 가장 양호한 것으로 판단된다. MACD 전략의 상승확률(up hit ratio)은 20일, 60일, 120일, 240일 각각 58.8%, 55.1%, 54.3%, 54.1%를 보이면서 앞서 실행한 모든 전략 중에서 가장 상승확률이 낮은 것으로 나타났다.

4.3 시스템 리스크 지표 실험 결과

최적화 결과 목적함수인 코스피(KOSPI) 지수의 표준편차를 극대화하는 주성분의 갯수(K)는 <표 12>와 같이 3개로 나타났다. 즉 주식의 섹터에 대한 정보를 3개의 주성분으로 압축할 때 목적함수가 최대화된다는 것이다.

<표 12> AR 전략 파라미터 최적화 결과

q1	q2	Sigma	K_rolling	K	Target Function
2	241	1.05	455	3	1043.548

<표 13>은 3.3 시스템 리스크 실험 설계에서 정의하였던 전략의 결과를 표로 정리한 것이다. 먼저 방어적 전략인 DOWN과 DECREASE 전략의 경우, 시그널이 발생하였을 경우 수익률을 무위험 수익률(투자 지양)로 인식하고 시그널이 발생하지 않을 경우 수익률을 시장지수(KOSPI) 수익률로 인식하였다. 반면 공격적 전략이 UP과 ENLARGE의 경우, 시그널이 발생했을 경우 시장지수(KOSPI) 수익률로 인식하고, 시그널이 발생하지 않을 무위험수익률로

인식하였다. 전반적으로 이동평균선 보조지표(ma5-ma20)의 부호를 기준으로 하는 UP과 DOWN 전략이 거리를 기준으로 하는 ENLARGE와 DECREASE 전략보다 더 유효한 전략으로 확인되었다. 샤프지수(Sharp ratio)를 기준으로 부호를 기준으로 하는 전략의 경우 하방 방어 목적의 DOWN 전략이 상향 조정 목적의 UP 전략에 비해 더 좋은 결과를 보였다.

5. 결론 및 시사점

본 연구는 인공지능 시대에 관심이 높아진 로보어드바이저에 사용될 수 있는 금융지표를 정의하였고, 파라미터 최적화를 통해 KOSPI 개별주식들에 대한 시장수익률을 초과하는 수익률을 달성할 수 있는 지에 대해서 분석사례이다. 본 연구에서는 목표로 하는 확률 분포(각 금융 지표를 구성하는 확률 변수)를 특정하기 어렵다는 문제를 마르코프 연쇄 몬테카를로(MCMC) 방식을 활용하여 표본을 추출하는 방식으로 해결하였다. 최적화는 각기 전략의 목적에 맞는 목적함수를 지정하고, 그 값을 최대화하는 파라미터를 구하는 방법을 활용하였다.

<표 13> 시스템 리스크(AR) 전략 간 비교

	KOSPI	60/40	DOWN	DECREASE	UP	ENLARGE
2005	28.25622	16.64634	27.3378	28.25622	19.61626	19.83271
2006	3.993852	3.597753	3.993852	3.993852	3.597753	3.597753
2007	32.25395	19.87517	27.95208	17.20071	27.30448	22.66158
2008	-40.7278	-25.8652	-19.1945	-28.4265	-27.8886	-33.7915
2009	49.65006	28.28066	45.92348	45.67009	29.46423	29.06677
2010	21.88237	13.42627	17.60889	17.82424	14.87907	14.71058
2011	-10.9829	-5.88083	-16.6263	-10.6039	-4.80324	-4.11563
2012	9.383045	6.463229	16.02705	13.77313	5.246977	5.053803
2013	0.715555	1.161755	0.715555	0.715555	1.161755	1.161755
2014	-4.76051	-2.24017	-0.15321	-1.37742	-2.61521	-2.61967
2015	2.386732	1.908635	2.386732	1.642499	3.280609	3.037872
2016	3.321759	2.38784	1.488579	-1.50955	3.338239	2.106134
2017	1.118206	0.677644	1.118206	1.118206	0.677644	0.677644
mean	7.422344	4.649163	8.35217	6.790549	5.635382	4.721521
Standard deviation	22.34398	13.37184	18.18477	18.33276	14.94123	15.54022
Sharp Ratio	0.332185	0.347683	0.459295	0.370405	0.37717	0.303826

연구 결과 기본적 지표인 PBR과 PER에 대한 전략은 동일한 조건의 시장수익률 보다 약 3~5배 정도의 수익률을 보이며 유효한 결과를 도출하였다. 최적 보유기간은 PBR, PER 각각 240일과 120일로, 이는 기본적 분석의 경우 장기적인 투자가 단기적인 투자보다 더 긍정적이라는 것을 의미한다. 또한 기본적 지표를 활용한 전략의 경우, 기술적 지표를 활용한 전략보다 개별 종목의 수익률이나 샤프 지수가 더 높은 경향을 보였다. 이는 기존의 기술적 지표를 활용한 기계학습 연구에 기본적 지표를 함께 사용한다면 더욱 우수한 결과를 기대할 수 있다는 가능성을 시사한다.

그리고, 기술적 지표인 볼린저밴드와 MACD는 가격 파생지표라는 점에서 기본적 지표보다 정보량이 부족하다는 한계를 지닌다. 이에 대해서 본 연구에서는 최대한 다양한 가정과 조건을 부여하여 다양한 전략을 수립하였고 그 중에서 가장 우수한 결과를 보인 전략들을 도출하였다. 실험 결과, 볼린저밴드를 활용한 전략은 시장 수익률을 압도할 뿐만 아니라, 절대적 수치로도 기본적 지표에 뒤지지 않는 수준의 결과를 도출하였다. 또한 상승확률 또한 모든 전략을 통틀어 가장 높은 수치를 기록하여 긍정적인 가능성을 보여 주었다. 반면에 MACD를 활용한 전략은 수익률이 높지 않았고, 상승확률도 저조한 것으로 나타났다.

마지막으로 시스템 리스크 지표는 변동성만으로 위험요소를 측정하는 기존의 방법의 한계점을 극복하기 위해 시도해 보았다. 주식 섹터에 대한 주성분 분석을 통해 위험이 집중되어 있는 정도를 계산하였다. 앞선 두 가지 지표와 달리 AR은 코스피 시장 전체의 위험을 측정하는 지표이고 이를 활용하기 위해 이동평균선을 함께 사용하는 4가지 전략을 수립하였다. 주목할 점은 이 지표가 위험지표인 만큼, 공격적인 전략뿐만 아니라 방어적인 전략을 고안하는 것에도 사용될 수 있다는 것이다. 실제로도 시장 전체의 하락 신호가 발생하였을 때 무익험 수익률에 투자하는 방어전략이 가장 높은 수익률을 보임으로서 방어적 전략의 유효성을 증명하였다.

본 연구를 통해 금융 지표에서 시장 수익률을 상회하는 매매전략을 도출할 수 있다는 것을 확인하였다. 향후 로보어드바이저를 개발하는 과정에서도 이번 실험에서 검증한 기본적인 전략에서 나아가 보다 정교하고 복합적인 전략을 도출하여 활용하는 것이 가능할 것이다. 추가로 이번 연구는 금융적 지표를 활용한 전략이 실제로 작동한다는 것으로 현재 시점에서 과거 시점의 정보만을 사용하더라도 유의미한 결과를 얻을 수 있다는 것을 시사한다. 하지만, 그 과정에서 파라미터 값을 과거 데이터를 활용하여 최적화시켰다는 점은 한계가 될 수 있다. 왜냐하면 이러한 실험은 과거의 움직임과 미래의 움직임이 일관되게 움직일 것이라는 가정을 내포하고 있기 때문이다. 즉, 활용한 전략이 과거 데이터에 과적합(overfitting)되어 미래에 대한 예측력이 떨어질 수 있다는 위험이 존재하는 것이다. 물론 이는 기계학습을 활용한 모든 연구에서도 발생할 수 있는 문제이다. 위 문제를 해결하기 위해서 앞으로의 연구에서는 부트스트랩(bootstrap)이나 배깅(bagging) 방법론을 통해 파라미터의 과적합을 방지하여 보다 우월한 모형을 제시할 가능성을 연구할 수 있을 것이다.

참고문헌

- 김동영, 박제원, 최재현, “SNS와 뉴스기사의 감성분석과 기계학습을 이용한 주가예측 모형 비교 연구”, *한국IT서비스학회지*, 제13권, 제3호, 2014, 221-233.
- 정현중, “시스템 리스크와 동적 자산배분”, 한국투자증권리포트, 한국투자증권, 2016.
- 김선웅, 안현철, “Support Vector Machines와 유전자 알고리즘을 이용한 지능형 트레이딩 시스템 개발”, *지능정보연구*, 제16권, 제1호, 2010, 71-92.
- 이우식, “딥러닝분석과 기술적 분석 지표를 이용한 한국 코스피주가지수 방향성 예측”, *한국데이터정보과학회지*, 제28권, 제2호, 2017, 287-295.

- 박재연, 유재필, 신현준, “기술적 지표와 기계학습을 이용한 KOSPI 주가지수 예측”, *정보화연구*, 제13권, 제2호, 2016, 331-340.
- Appel, G., *Technical analysis : power tools for active investors*, FT Press, 2005.
- Chib, S. and E. Greenberg, “Understanding the metropolis-hastings algorithm”, *The American Statistician*, Vol.49, No.4, 1995, 327-335.
- Domingos, P., “A Few Useful Things to Know about Machine Learning”, *Communications of the ACM*, Vol.55, No.10, 2012, 78-87.
- Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, “Introducing markov chain monte carlo”, *Markov chain Monte Carlo in practice*, Vol.1, 1996, 19.
- Kabasinskas, A. and U. Macys, “Calibration of Bollinger Bands Parameters for Trading Strategy Development in the Baltic Stock Market”, *Engineering Economics*, Vol.21, No.3, 2010, 244-254.
- Kritzman et al., “Principal components as a measure of systemic risk”, *The Journal of Portfolio Management*, Vol.37, No.4, 2011, 112-126.
- Mahendran, N., Z. Wang, F. Hamze, and N. de Freitas., “Adaptive MCMC with Bayesian Optimization”, Technical Report, Statistics Department, University of British Columbia, 2012.
- Chang, P.H.K. and C.L. Osler, “METHODICAL MADNESS : Technical Analysis and The Irrationality of Exchange-rate Forecasts”, *The Economic Journal*, Vol.109, No.458, 1999, 636-661.
- Bernard, V.L. and J.K. Thomas, “Post-Earnings-Announcement Drift : Delayed Price Response or Risk Premium?”, *Journal of Accounting Research*, Vol.27, 1989, 1-36.
- Lo, A.W., H. Mamasky, and J. Wang, “Foundations of Technical Analysis : Computational Algorithms, Statistical Inference, and Empirical Implementation”, *Journal of Finance*, Vol.4, No.4, 2000, 1705-1765.
- Yildirim, I., “Bayesian Inference : Metropolis-Hastings Sampling”, Technical Report, Department of Brain and Cognitive Sciences, University of Rochester, 2012.

◆ About the Authors ◆



강민철 (danielkang357@gmail.com)

한양대학교 경영학부와 컴퓨터공학 학사 수여 후 일본 E-Commerce 기업에서 데이터사이언티스트로 근무 하였다. 현재 맥킨지 재팬에서 데이터사이언티스트로 근무 중이다. 인공지능과 머신러닝 기법을 통한 주식 알고리즘 개발 등에 관심이 많다.



임규건 (gglim@hanyang.ac.kr)

한양대학교 경영대학 임규건 교수는 KAIST 전산학 학사, POSTECH 컴퓨터 석사, KAIST 경영공학 박사학위를 취득하였고, 삼성전자, KT, 국제전자상거래연구센터(ICEC) 연구위원, 세종대학교 경영학과 교수를 역임하였다. 관심 분야는 혁신 비즈니스모델, IT서비스 혁신, 인공지능과 경영, e-Business 등이며, 2018년 IT서비스 우수연구인상을, 2009년 IT Innovation 유공자 지식경제부 장관 표창과 2007년 SW산업발전 유공자 정통부 장관 표창을 수여하였다. 주요 저서로는 '경영을 위한 정보기술', 'e-비즈니스 경영', '디지털경제시대의 경영정보시스템' 등 전문서적과 다수의 논문과 특허가 있다. 또한, 아시아최초 상용인터넷인 KORNET상용화, 중국 Shanghai Telecom SI사업전략, 한국 영화기술 로드맵, KTI 사업전략, 나라장터(G2B) 효과평가, 행정정보화(G4C) 성과분석, 국가정보보호지수개발, 국방정보화 수준평가모형, IT혁신인력양성종합대책, 국가디지털식별체계(UCI) 개선방안, 저작권정품인증제도, SW사업자신고제도개선, SW기술자신고제도개선 등 다양한 IT혁신 분야의 프로젝트를 수행하였다.