

기계학습을 활용한 상품자산 투자모델에 관한 연구

송진호

국민대학교 비즈니스IT전문대학원
(safe1004@gmail.com)

최흥식

국민대학교 비즈니스IT전문대학원
(hschoi@kookmin.ac.kr)

김선웅

국민대학교 비즈니스IT전문대학원
(swkim@kookmin.ac.kr)

.....

상품자산(Commodity Asset)은 주식, 채권과 같은 전통자산의 포트폴리오의 안정성을 높이기 위한 대체투자 자산으로 자산배분의 형태로 투자되고 있지만 주식이나 채권 자산에 비해 자산배분에 대한 모델이나 투자전략에 대한 연구가 부족한 실정이다. 최근 발전한 기계학습(Machine Learning) 연구는 증권시장의 투자부분에서 적극적으로 활용되고 있는데, 기존 투자모델의 한계점을 개선하는 좋은 성과를 나타내고 있다.

본 연구는 이러한 기계학습의 한 기법인 SVM(Support Vector Machine)을 이용하여 상품자산에 투자하는 모델을 제안하고자 한다. 기계학습을 활용한 상품자산에 관한 기존 연구는 주로 상품가격의 예측을 목적으로 수행되었고 상품을 투자자산으로 자산배분에 관한 연구는 찾기 힘들었다.

SVM을 통한 예측대상은 투자 가능한 대표적인 4개의 상품지수(Commodity Index)인 골드만삭스 상품지수, 다우존스 UBS 상품지수, 톰슨로이터 CRB상품지수, 로저스 인터내셔널 상품지수와 대표적인 상품선물(Commodity Futures)로 구성된 포트폴리오 그리고 개별 상품선물이다. 개별상품은 에너지, 농산물, 금속 상품에서 대표적인 상품인 원유와 천연가스, 옥수수과 밀, 금과 은을 이용하였다.

상품자산은 전반적인 경제활동 영역에 영향을 받기 때문에 거시경제지표를 통하여 투자모델을 설정하였다. 주가지수, 무역지표, 고용지표, 경기선행지표 등 19가지의 경제지표를 이용하여 상품지수와 상품선물의 등락을 예측하여 투자성과를 예측하는 연구를 수행한 결과, 투자모델을 활용하여 상품선물을 리밸런싱(Rebalancing)하는 포트폴리오가 가장 우수한 성과를 나타냈다. 또한, 기존의 대표적인 상품지수에 투자하는 것 보다 상품선물로 구성된 포트폴리오에 투자하는 것이 우수한 성과를 얻었으며 상품선물 중에서도 에너지 섹터의 선물을 제외한 포트폴리오의 성과가 더 향상된 성과를 나타남을 증명하였다.

본 연구에서는 포트폴리오 성과 향상을 위해 기존에 널리 알려진 전통적 주식, 채권, 현금 포트폴리오에 상품자산을 배분하고자 할 때 투자대상은 상품지수에 투자하는 것이 아닌 개별 상품선물을 선정하여 자체적 상품선물 포트폴리오를 구성하고 그 방법으로는 기간마다 강세가 예측되는 개별 선물만을 골라서 포트폴리오를 재구성하는 것이 효과적인 투자모델이라는 것을 제안한다.

주제어 : 상품자산, 기계학습, SVM, 투자모델

.....

논문접수일 : 2017년 10월 11일 논문수정일 : 2017년 11월 22일 게재확정일 : 2017년 12월 1일
원고유형 : 일반논문 교신저자 : 최흥식

1. 서론

인공지능(Artificial Intelligence)을 이용한 서비스가 생활 속에 등장하기 시작했다. 인공지능 냉장고, 인공지능 스피커 등 가전과 통신에서는 인

공지능 기술을 제품에 적용하는 것이 보편화 되고 있다. 금융분야에서도 골드만 삭스(Goldman Sachs)는 당사가 투자한 켄쇼(Kensho)의 인공지능 기술을 이용하여 600명의 주식 트레이더가 처리하던 업무를 2명이 처리할 수 있도록 프

세스를 개선하였고 15명이 4주간 처리할 분석업무를 5분만에 처리할 수 있다는 사례가 보도되고 있다. 특히 인공지능 분야 중 기계학습을 통한 빅데이터 분석이 금융산업 전반에서 활발하게 적용되고 있다.

매매체결 및 결제업무와 같은 백오피스(Back Office) 업무에서부터 인공지능 기술의 적용이 시작되어 이제는 투자자의사결정과 같은 프론트오피스(Front Office)업무에 있어서도 인공지능 기술을 이용하여 새로운 시각으로 시장을 분석하고 예측하는 시도가 활발하다. 금융시장에는 수많은 데이터들이 존재하며 매일 다양한 경제 뉴스와 각종 경제지표들이 발표되고 개별기업과 산업에 대한 뉴스와 공시들이 쏟아져 나오기 때문에 다량의 정보를 처리할 수 있는 기계학습 기술에 대한 관심이 크다. 최근 우리나라에서도 ‘로보어드바이저’라는 기계학습을 바탕으로 한 알고리즘 모델이 주식투자자와 자산배분에서 기존의 자산운용사나 투자자문사와 경쟁을 시작하였다.

기계학습이론을 통한 증권시장 분석 및 투자모델 설정에 관한 연구도 매우 활발하게 이루어지고 있다. 기존의 재무연구에서 사용했던 선형(Linearity) 시계열 분석들이 가지는 한계를 인공지능 예측모형 등 기계학습 이론을 활용하여 연구를 개선시키고 발전시키고 있다. 이러한 기계학습을 통한 금융시장 분석은 주식시장에서 활발하게 이루어지고 있는데 과거의 증권시장관련 수치기반의 계량적 금융데이터를 학습시켜 향후의 주가지수 혹은 주식의 움직임을 예측하는 모델을 만드는 연구와 주식시장에 관련된 다양한 뉴스나 코멘트 등 방대한 텍스트자료를 기반으로 학습시켜 향후의 시장 혹은 기업의 주가 방향을 예측하는 연구들이 다양한 방법으로 행해지

고 있다.

금융데이터를 이용한 기계학습 연구는 초기에는 주가나 주가지수 예측에 관한 연구가 활발했으며(Heo and Yang, 2015), 최근에는 투자 성과를 높이기 위한 지능형 매매시스템에 관한 연구와(Kim and Ahn, 2010; Choi et al., 2011), 변동성에 관한 연구(Ra et al., 2016; Kim and Choi, 2017) 등 파생상품 시장에서도 유용한 연구들이 늘어나고 있다.

‘뉴스와 주가’라는 연구주제는 빅데이터를 처리할 수 있는 기계학습으로 연구가 가능하게 되었는데 최근에는 트위터와 페이스북 등 다양한 SNS를 통해 얻어지는 텍스트 기반으로 한 분석도 발표되고 있다. 뉴스의 감성분석 결과를 이용하여 주가지수 변동을 예측한 투자모델에 관한 연구(Kim et al., 2012)와 감성사전을 이용한 감성분석과 기술적 지표를 결합하여 주가지수를 예측하는 연구(Cha and Hong, 2016) 등 시장에 대한 감성적 분석을 통한 시장예측도 기계학습을 이용하여 다양하게 발전하고 있다.

금융투자시장은 증권시장이 가장 큰 시장이지만 리츠(Real Estate Investment Trusts, REITs)와 같이 금융 상품화된 부동산투자 상품, 금, 옥수수, 원유와 같은 상품투자 상품 및 다양한 대체투자 시장이 존재하고 있으며 기관투자자를 비롯한 대규모의 자산의 투자는 이러한 다양한 금융상품의 포트폴리오로 구성되어 있다. 따라서 기계학습을 이용한 투자자산 분석과 투자모델 구축은 다양한 투자자산에 적용될 수 있으며 본 연구는 이러한 시도를 상품자산에 적용해 보려고 한다.

상품자산에 대한 투자는 대표적 투자자산인 주식, 채권 및 부동산 등과 상관관계가 낮거나 역의 상관관계를 나타냄으로 투자 포트폴리오의

위험을 낮추고 수익률을 제고시키는 효과가 있다는 CAPM(Capital Asset Pricing Model)이론의 공동 창시자인 존 린트너(John Lintner)박사의 연구 이래로 꾸준히 검토되어 왔다(Lintner, 1983). 상품투자에 대한 전반적인 분석이 이루어진 연구논문(Gorton and Rouwenhorst, 2006; Bhardwaj et al., 2015)에서는 월별수익률을 이용하여 상품 선물과 주식, 채권의 상관관계가 역의 상관관계를 나타내며 분산투자 효과가 있음을 알려주었다. 2015년의 논문에서는 이러한 상관관계가 유지되었지만 금융위기 시기에는 상품자산과 다른 자산과 상관관계가 일시적으로 높아지는데 이는 경기 사이클에서 나타나는 현상과 유사하다고 분석했다. 또한 이 두 논문에서 상품투자가 인플레이션 헷지에 유용하다는 일반적인 상식도 분석했는데 주식 및 채권은 인플레이션과 역의 상관관계를 나타냈고 상품선물 인덱스는 양의 상관관계를 나타냈다. 인플레이션과 상관관계분석은 기간이 길수록 더욱 높은 상관관계를 나타내는데 이는 인플레이션 헷지 수단으로 상품투자가 유용하다는 의미를 시사한다.

일반적으로 상품자산에 대한 투자는 해외에서는 상품 포트폴리오로 이루어진 상품지수 ETF (Exchange-Traded Fund)에 대한 투자, 상품선물에 대한 직접 투자 및 상품자산에만 투자하는 CTA(Commodity Trading Advisor)를 통한 간접 투자가 보편적이다.

CTA 혹은 관리형 선물(Managed Futures)투자에 대해서는 시카고상업거래소(Chicago Mercantile Exchange, CME)가 제시한 자료에 따르면 주식, 채권, 관리형 선물로 포트폴리오를 구성했을 때 총투자 중 20%를 관리형 선물에 투자했을 때 포트폴리오 분산성을 개선하여 효율적 투자성을 얻을 수 있다고 분석했다(CME Group, 2012). 한

편으로는 CTA들의 성과가 다소 상향으로 과장된 성향이 있으며 CTA들이 실제 성과를 창출하는 것에도 의문이 있다는 의견을 제시하는 연구(Bhardwaj et al., 2014)도 있다. 국내에도 2008년 금융위기 이후에 전통자산에 대한 대체상품으로 CTA에 대한 투자가 소개되어 유명 CTA펀드를 재간접투자펀드(Fund of Funds)로 구성하여 시장에 출시하였으나 단기성과 부진으로 현재 시장에서 찾아보기가 힘들다.

최근 다양한 상품ETF 등장, 원유 및 다양한 상품을 기초자산으로 하는 파생결합증권(Derivative Linked Securities, DLS)상품 및 해외상품선물에 대한 직접투자 등 투자자들의 상품자산에 대한 많은 관심에 부응하는 투자상품들이 등장했다. 하지만 포트폴리오의 효율성 증대를 위한 분산투자 이외에 특별한 투자전략에 대한 연구는 찾아보기 힘들다. 주요 증권사에서 자산배분에 대한 가이드라인을 제시하면서 상품투자에 대한 비중 정도를 제시하고 있는 정도이고 자산운용사는 상품투자에 관련한 상품을 출시하는 경우 그 상품투자에 대한 설명을 제시할 뿐 상품투자 전략에 대한 전반적인 연구가 부족한 상황이다.

상품자산 분야에 기계학습을 이용한 연구들은 대부분 개별 상품자산 가격을 예측하는 연구들이었다. 기계학습을 활용하여 기존 시계열 분석 통계기법인 ARIMA(Auto Regressive Integrated Moving Average)보다 유가의 예측성과를 높인 연구(Xie et al., 2006)와 유가 예측을 위해 SSL (Semi-Supervised Learning) 기법을 활용하여 시계열 데이터 자체의 잡음을 해결하고 예측력을 높인 연구 등이 있었다(Park et al., 2011). 최근에는 기계학습 이론을 활용하여 상품 가격을 통해 말레이시아의 환율을 예측하려는 연구 등 상품 가격을 활용한 거시지표 예측연구도 시도되고

있다(Ramakrishnan et al., 2017). 하지만 아직 자산배분 차원에서 상품자산에 대한 접근은 찾아보기 힘들다. 본 연구는 최근 발전한 기계학습 모형을 기반으로 자산배분의 관점에서 상품투자 전략에 대한 전략을 연구해 보고자 한다.

2. 연구의 방법론

본 연구에서는 SVM(Support Vector Machine)은 기계학습의 한 분야로서 주로 분류를 위해 사용한다. 다른 두 집단에 대한 데이터 정보를 알고 있을 때, SVM은 주어진 데이터 집합의 정보를 바탕으로 하여 새로운 데이터가 어느 집단으로 분류되는지를 판단해준다. SVM은 두 집단의 데이터를 기하학적인 벡터 공간에 사상(Mapping)시킨 후 적당한 분류 기준을 공간 위에 표현해 준다. 이후 새로운 데이터의 위치에 따라 그 데이터가 어느 집단에 속할지를 분류 기준에 의해 판단해준다. SVM은 기계학습 중에서 인공신경망과 더불어 예측력이 높다고 알려져 있으며 여러 분야에서 SVM을 이용한 연구를 진행하고 있다(Park and Hansen, 2012; Li et al., 2008).

2.1 선형 SVM

기초적인 SVM은 직선을 분류 기준으로 사용하는 선형 SVM이다. 오분류(Miss-classification) 없이 선형방정식(Linear Equation)으로 데이터가 완전하게 분리되는 경우이다. 이때 각 데이터 집단에서 선형식과 가장 가깝게 있는 데이터들을 서포트 벡터(Support Vectors)라고 하며, 두 집단의 서포트 벡터들 사이의 거리를 마진(Margin)이라고 한다. SVM 모델에서는 마진이 최대가 되도록

선형식이 결정되며 이를 최대 마진 초평면(Maximum-margin Hyperplane)이라고 한다. 최대 마진 초평면은 새로운 데이터를 분류해주는 기준이 된다(Flach, 2012).

한편 오분류가 없이 데이터를 완전하게 선형식으로 분리할 수 없는 경우에는 오분류를 용인하되 그 정도를 최소화 하도록 선형식이 결정된다. 이때 SVM 모델은 ‘마진을 최대화’, ‘오분류 정도를 최소화’ 이 두 가지의 목적을 고려하며 선형식을 결정한다. 사용자는 모델 적용 시 Cost 라는 매개변수(Parameter)의 값을 통해 두 목적 중 어느 곳에 더 비중을 둘 지를 정할 수 있다(Oh, 2008).

2.2 비선형 SVM

SVM 모델의 최대 마진 초평면을 구하는 과정에 커널 방법(Kernel Method)을 적용하는 비선형 SVM이 제안되었다(Boser et al., 1992). 이의 핵심은 ‘고차원 공간으로의 데이터 이동’이다. 이동된 고차원 공간에서의 초평면은 기존 공간에서는 선형이 아니므로 ‘비선형 SVM’이란 용어가 사용된다. 고차원 공간으로 데이터 이동이 분류 성과를 더 높여 주기도 한다. 하지만 실제로는 분류 성과를 높여주는 고차원 공간으로의 이동을 가능하게 해주는 함수를 찾는 것은 불가능하다. 다만 고차원 공간으로의 데이터 이동을 위해서 커널 함수(Kernel Function)를 이용한다. 커널 함수는 비선형 SVM을 가능케 해주며 또한 여러 종류의 커널 함수의 선택은 모델의 성과를 결정시켜주는 중요한 요소가 될 수 있다(Tay and Cao, 2002).

현재 많은 프로그래밍 언어에서 SVM의 구현이 가능하고, R Programming에서는 ‘e1071’,

‘kernlab’ 등의 패키지를 이용하여 SVM 모델링과 분석을 할 수 있다.

3. SVM 모델링과 상품편입 포트폴리오 구성 아이디어

SVM 모델을 통한 실제 분석을 원하는 사용자들을 위해 R, MATLAB, Python, Java, C/C++과 같은 프로그래밍 언어는 SVM 패키지를 무료로 제공한다. 사용자는 프로그래밍 위에 원하는 데이터를 불러온 후 직접 SVM 모델링을 할 수 있다. 본 연구에서는 R의 SVM 패키지를 통해 모델링을 구현한다(Lantz, 2013).

예측 대상은 상품지수 및 상품선물의 월별 등락이다. 상품지수는 상품투자에 대한 벤치마크로 사용하는 대표적인 4개의 지수를 선정하였고 상품선물은 에너지, 농산물, 금속 세 섹터에서 유동성과 인지도가 높은 대표적인 상품 각 두가지를 선택하였다. 예측대상 자료는 블룸버그(Bloomberg) 자료를 사용하였다.

골드만삭스 상품지수(Goldman Sachs Commodity Index)는 현재 S&P(Standard & Poor's)에서 관리 발표되고 있는데 24개의 상품을 기초자산으로 구성하고 있으며 에너지 섹터의 비중이 높은 특성이 있다. 다우존스 UBS 상품지수(Dow Jones UBS Commodity Index)는 블룸버그(Bloomberg)에서 관리 발표하고 있으며 22개의 상품을 기초자산으로 구성하고 있으며 7개 섹터로 구분되어 상품간 비중을 균형 있게 유지하는 특성이 있다. 톰슨 로이터 코어상품 CRB 지수(Thomson Reuter/Core Commodity Index)는 가장 오래된 상품지수로 기존에는 28개의 상품

을 기초자산으로 구성하고 있었으나 최근 19개의 상품을 기초자산으로 구성하고 있으며, 원래 미국 노동통계국에서 발표를 시작하였으나 현재는 톰슨로이터사에서 관리 발표하고 있다. 로저스 인터내셔널 상품지수(Rogers International Commodity Index)는 유명한 상품투자자인 짐 로저스가 설계한 지수로 38개의 상품을 기초자산으로 구성하고 있는데, 전세계 거래소에서 가장 활발하게 거래되는 유동성에 초점을 맞추어 기초자산을 선택하는 특성이 있다.

상품선물은 총 6개 상품으로 에너지 섹터에서 크루드 오일과 천연가스, 농산물 섹터에서 옥수수과 밀, 금속 섹터에서 금과 은을 예측대상 상품선물로 선택하였다. 각 상품들은 해당 섹터에서 가장 활발하게 거래되는 상품 중에 일상 생활에서도 체감할 수 있는 기초적인 상품들을 기준으로 선정하였다. 모든 상품들은 CME에 상장되어 있고 CME의 GLOBEX시스템을 통해 24시간 활발하게 거래되고 있다.

<Table 1>에서 Output Data에 대한 정보를 확인할 수 있다. 모델링을 위해 상품선물의 가격 변화에 영향을 줄 수 있다고 판단한 19가지 Input Data의 월 단위 값들을 수집하고, 차월 상품선물의 가격변화의 값을 구한 후, (+)부호 혹은 (-)부호에 맞춰 각각 ‘Up’, ‘Down’으로 이름 지어(Labeling) Output 값을 할당하였다. 데이터의 총 기간은 1990년 1월~ 2017년 5월까지 총 320개의 월별 데이터이다. 그 후 전체 데이터 중 1990년 1월 ~ 2008년 12월까지의 195개 데이터를 훈련데이터(Training Data)로 설정하고, 2009년 1월 ~ 2017년 5월까지의 125개 데이터를 실험데이터(Test Data)로 설정하여 훈련데이터의 Input 변수와 Output 변수 간의 인과관계를 바탕으로 SVM 모델링을 실행하고 실험데이터로 이

〈Table 1〉 Output Data

Name	Official Website
Goldman Sachs Commodity Index	kr.spindices.com/performance-overview/commodities/sp-gsci
Dow Jones UBS Commodity Index	www.bloombergindeces.com/bloomberg-commodity-index-family/
Thomson Reuters/ Core Commodity CRB index	financial.thomsonreuters.com/en/products/data-analytics/market-data/indices/commodity-index.html
Rogers International Commodity Index	www.rogersrawmaterials.com/home.asp
WTI Crude Oil Futures	www.cmegroup.com/trading/energy/crude-oil/light-sweet-crude_contractSpecs_futures.html
Natural Gas Futures	www.cmegroup.com/trading/energy/natural-gas/natural-gas_contract_specifications.html
Corn Futures	www.cmegroup.com/trading/agricultural/grain-and-oilseed/corn_contract_specifications.html
Wheat Futures	www.cmegroup.com/trading/agricultural/grain-and-oilseed/wheat_contract_specifications.html
Gold Futures	www.cmegroup.com/trading/metals/precious/gold_contractSpecs_futures.html
Silver Futures	www.cmegroup.com/trading/metals/precious/silver_contractSpecs_futures.html

를 검증하였다. 모델링에 사용한 변수는 커널 함수를 제외하고는 모두 기본값(Default)을 사용하였다. 커널 함수는 선형 SVM을 위한 Linear, 비선형 SVM을 위한 Radial basis 두 가지를 사용하였다.

3.1 Input Data와 선정 이유

<Table 2>에 설명된 Input Data는 19개 경제지표이다. Input Data도 월간 등락률을 사용했다. 상품투자는 경기변화와 가장 민감한 영향을 가지고 있기에 세계경기를 가늠할 수 있는 경제지표들을 선택했다. 전세계적으로 상품자산은 달러로 결제되므로 우선 미국 달러 인덱스를 선정하였고 경기변화의 선행성을 띠는 주가지수(미국, 한국), 세계경기의 움직임을 파악하기 위하여 미국 경기선행지수 및 그 구성항목을 이루는

미국의 경제지표들과 세계경기를 이끌고 있는 또 하나의 축인 중국의 무역지수와 소비자물가지수, 그리고 자국의 자원 없이 무역으로 경제활동을 하고 있는 한국의 수출지표를 선택하였다.

간단하게 19개 변수를 살펴보면 첫 번째, 달러 인덱스는 유로, 엔, 파운드, 캐나다 달러, 스웨덴 크로네, 스위스 프랑 등 경제규모가 크거나 통화가치가 안정적인 통화를 기준으로 미 달러화 가치를 지수화한 것이다. 두 번째, S&P 500은 뉴욕거래소의 대표적인 500개 기업 주가의 움직임을 하나의 지표로 나타낸 것이다. 세 번째, KOSPI200은 한국거래소의 대표적인 200개 기업 주가의 움직임을 하나의 지표로 나타낸 것이다. 네 번째, 한국수출은 대한민국의 상품무역의 거래량을 나타낸다. 다섯 번째, 미국 총 노동력 실업률은 가구 노동력을 기반으로 실업자수를 추측하는 지표이다. 여섯 번째, 미국 고용인구비율

은 전체 취업연령인구 대비 취업인구의 비율을 의미하며 취업비율을 나타낸다. 일곱 번째, 미국 신규 개인주택착공 구조별 합계는 일정기간 동안 시작된 신규주택 혹은 건물의 수를 추적하여 발표하는 지표이다. 여덟 번째, 미국 제조 및 재고 합계는 종종 사업 재고로 언급되는데, 제조업 및 도소매 무역업에서의 재고량(생산자가 여전히 보유하고 있는 판매 또는 출하가 즉시 가능한 상품, 또는 추가처리 없이 재판매목적으로 취득한 상품)을 의미한다. 아홉 번째, 미국 민간 비농

업부문 고용자 합계 전월대비 순변동 계절조정 값이다. 열 번째, 미국 경제활동 참가율은 취업연령 인구 대비 퍼센트로 나타낸 총 노동인구이다. 열 한 번째, 미국 산업생산은 광업 및 채석업, 제조업, 공공 유틸리티(전기, 가스, 수도 공급)산업에 소속된 업체의 산출량을 측정하여 이를 기준으로 발표하는 지표이다. 열 두 번째, 미국 개인 소비지출은 소비자가 구매하는 재화 및 용역의 전반적인 가격변동을 추적하여 산출하는 것이다. 열세 번째, 컨퍼런스보드 미국 선행지수는

〈Table 2〉 Input Data

	Name	Source
1	U.S. Dollar Index	US Federal Reserve Board (FRB)
2	S&P 500	Standard & Poors
3	KOSPI 200	Korea Exchange (KRX)
4	South Korea Export Volume Index - Manufacturing Products	Bank of Korea (BOK)
5	U-3 US Unemployment Rate Total in Labor Force Seasonally Adjusted	US Bureau of Labor Statistics
6	US Employment Population Ratio Total in Labor Force SA	US Bureau of Labor Statistics
7	US New Privately Owned Housing Units Started by Structure Total SAAR	U.S. Census Bureau
8	US Manufacturing & Trade Inventories Total MoM SA	U.S. Census Bureau
9	US Employees on Nonfarm Payrolls Total Private MoM Net Change SA	US Bureau of Labor Statistics
10	US Labor Force Participation Rate SA	US Bureau of Labor Statistics
11	US Industrial Production SA (2002=100)	US Federal Reserve Board (FRB)
12	US Personal Consumption Expenditure Core Price Index YoY SA	US Bureau of Economic Analysis
13	Conference Board US Leading Index MoM	US Conference Board
14	US Capacity Utilization % of Total Capacity SA	US Federal Reserve Board (FRB)
15	MNI Chicago Business Barometer (SA)	Chicago National Association of Purchasing Management
16	Index of Institute for Supply Management: ISM	US Institute for Supply Management
17	US Initial Jobless Claims SA	US Department of Labor
18	China Imports & Exports Trade Balance Value	Customs General Administration P.R.China
19	China CPI YoY	National Bureau of Statistics of China

경제전반에 변화가 발생하기 이전에 미리 변동하는 경향이 있는 경제변수를 관찰하여 미래 경제상황을 예측하기 위한 지표를 말한다. 열 네 번째, 미국 총설비가동률 대비 설비가동률은 국가의 설치 생산용량이 상품 및 용역의 생산에 사용되는 정도를 추적하는 지표이다. 열다섯 번째, 미국 시카고 구매관리자지수와 열 여섯 번째 공급관리협회 제조업 구매관리자지수는 제조, 건설 및 서비스 기업의 구매담당 관리자들이 느끼는 체감도를 추적하는 지표로 전체 체감도 지수는 일반적으로 생산, 주문, 재고, 고용, 가격 등의 질문에 대한 답변 내용을 이용하여 산출한다. 열 일곱 번째, 미국 신규실업수당 청구건수는 특정 기간 동안 처음으로 해당 정부 노동기관에 실업수당을 신청한 근로자수를 나타내며 실업수당을 신규 수령하게 된 인구를 나타낸다. 열 여덟 번째, 중국무역수지는 중국의 상품 수출과 수입의 차이를 계산하여 무역수지의 값을 추적하는 지표이다. 마지막으로 열 아홉 번째, 중국 소비자물가지수는 중국의 소비자 상품 및 서비스의 시장 바스켓에 대하여 소비자가 지불하는 가격을 측정하여 산출하는 지표로, 이 수치의 성장률은 인플레이션율을 나타낸다.

3.2 모델링 결과를 활용한 성과분석 아이디어

모델링을 기반으로 한 상품편입 포트폴리오의 성과분석은 다음과 같이 3가지의 포트폴리오의 성과 비교분석을 통해 이루어 졌다. 첫 번째 포트폴리오는 벤치마크 지수로서 주식을 대표하는 S&P500지수, 현금을 대표하는 달러 인덱스 지수, 채권을 대표하는 미국국채10년물 지수를 각각 1/3 비중만큼 투자한 전통적 포트폴리오(Basic Portfolio, BasicPort), 그리고 두 번째 포트

폴리오는 벤치마크 지표인 전통적 포트폴리오에 매기간 조건 없이 상품선물을 포함하여 각 비중을 1/4로 한 모델을 기반하지 않은 포트폴리오(Non-model Portfolio, Non_MP), 마지막으로 세 번째 포트폴리오는 전통적 포트폴리오에 SVM 모델이 'Up'으로 예측하는 경우에만 다음 기간에 상품선물을 포함하게 되는 모델 기반의 포트폴리오(Model Portfolio, ModelPort)이다.

4. 분석결과

4장에서는 전통적 포트폴리오에 포함할 상품선물의 종류에 따라 각 절을 구분하였다. 4.1절에서 전통적 포트폴리오에 포함되는 상품선물은 개별상품이 아닌 상품선물 포트폴리오 형태이다. 대상은 시장에서 매매 가능한 대표적 상품선물지수 4가지, 그리고 개별선물을 이용하여 동일한 비중으로의 투자를 가정한 지수를 별도로 포함하였다. 4.2절에서는 전통적 포트폴리오에 개별선물 단 하나만을 포함하는 선물 포트폴리오에 대한 분석을 실시하였다. 4.2절을 통하여 전통적 포트폴리오에 편입함으로써 포트폴리오 성과를 향상시켜주는 개별상품의 종류를 선택할 수 있다. 4.3절에서 전통적 포트폴리오에 편입되는 선물의 형태는 SVM 모델을 바탕으로 사용자가 매 기간마다 새롭게 구성하는 리밸런싱 포트폴리오이다. 매 기간마다 Input Data를 통하여 개별선물의 등락을 예측하고 지수상승이 예측되는 개별선물만을 선택하여 선물 포트폴리오를 구성하여 이를 전통적 포트폴리오에 편입하여 성과 분석을 하였다. 4.4절에서는 4.3절의 내용을 유지하되, 훈련기간과 실험기간을 바꿔서 또 다른 시장상황에 대한 분석결과를 도출하였다.

4.1 대표적 상품지수와 개별선물의 동일가중 가격지수

첫째로 전통적 포트폴리오에 편입할 상품선물은 총 5가지로 대표적 상품지수 4가지 및 6개의 대표적 개별상품선물(WTI, 천연가스, 옥수수, 밀, 금, 은)의 동일가중 가격지수이다. 4가지의 대표적 상품지수는 각각 Goldman Sachs Commodity Index(GSCI), Dow Jones UBS Commodity Index(DJUI), Thomson Reuters/Core Commodity CRB Index(TRCI), Rogers International Commodity Index(RI)이다. 대표적 상품지수는 투자하기가 쉬울 뿐더러 여러 가지 종류의 상품선물을 포함하고 있기 때문에 상품에 대한 분산투자 효과를 얻을 수 있다는 장점이 있다. 이에 더해 6개의 개별선물 동일가중 가격지수는 각각의 개별상품선물의 가격을 바탕으로 동일가중 가격을 직접 계산하여 상품지수와 더불어 별도로 예측대상에 포함한 후 분석을 실시하였다.

본 연구에서의 총 데이터 기간과 실험데이터 기간, 테스트데이터 기간의 분류는 3.1절에서 언급한 기간 구성을 기본적으로 따르지만 본 절에서는 상품지수의 역사적 데이터가 부족하여 총 데이터 기간은 1999년 2월~2017년 5월이다. 따라서 1999년 2월 ~ 2008년 12월까지의 데이터를

훈련데이터로 설정하고, 2009년 1월 ~ 2017년 5월까지의 데이터를 실험데이터로 설정하였다. 훈련데이터를 바탕으로 모델링을 하여 실험데이터에 적용 및 예측을 실시하였고, 실험기간에 대한 예측수익률과 실제수익률을 비교하며 각 경우의 예측률을 <Table 3>와 같이 도출하였다.

<Table 3>는 상품지수에 대한 모델의 예측력은 커널 함수와 관계없이 50%를 넘지 못하고 있다. 반면에 6개 개별선물 동일가중 가격지수의 예측률은 커널 함수와 관계없이 53%임을 알 수 있다. 4.1절에서의 모델링은 커널 방법에 의한 성과 향상 효과를 거의 내지 못했다.

<Table 4>에는 모델이 Output을 ‘Up’으로 예측하는 경우에만 전통적 포트폴리오에 각 상품선물을 편입하고 그렇지 않은 경우에는 해당 기간에 전통적 포트폴리오만을 구성하는 포트폴리오의 누적수익률 값을 커널 함수별로 정리하였다. 또한 가시적인 분석결과 파악을 위해 GSCI에 대한 각 포트폴리오의 누적수익률 곡선과 6개 개별선물 동일가중 가격지수에 대한 각 포트폴리오의 누적수익률 곡선을 <Figure 1>에 나타냈다. 커널의 종류에 관계없이 누적수익률이 비슷했기 때문에 Linear 커널 함수에 대한 누적수익률 곡선만 나타냈다.

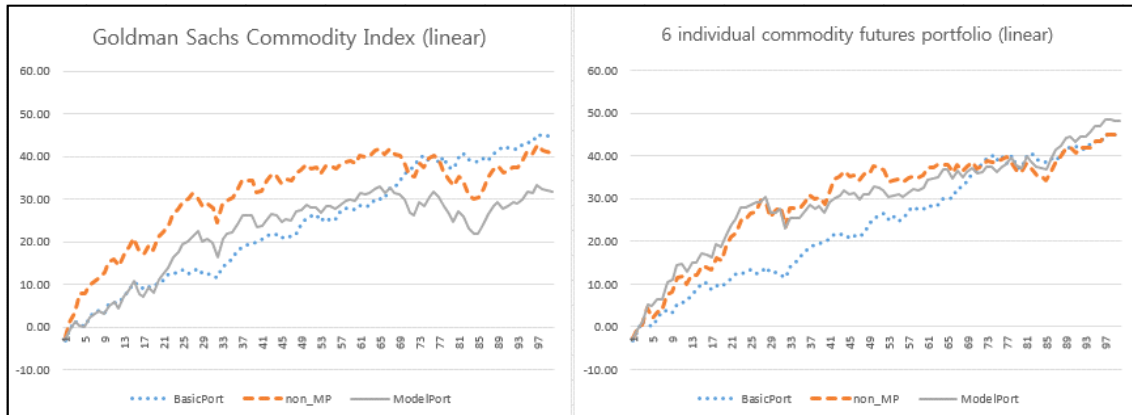
결과를 분석해보면 기본적으로 전통적 포트폴리오에 대표적 상품선물지수를 편입하였을

<Table 3> Accuracy of Representative Commodity Index and 6 Individual Commodity Futures Portfolio

Commodity Futures Index	Accuracy(Linear)	Accuracy(Radial)
GSCI	48%	49%
DJUI	46%	50%
TRCI	48%	48%
RI	48%	48%
6 Individual Commodity Futures Portfolio	53%	53%

〈Table 4〉 Cumulative Yield of Model Portfolio for Representative Commodity Index and 6 Individual Commodity Futures Portfolio

Commodity Index	Linear		Radial Basis	
	Portfolio	Cumulative Yield	Portfolio	Cumulative Yield
GSCI	BasicPort	44.76	BasicPort	44.76
	Non_MP	40.71	Non_MP	40.71
	ModelPort	31.59	ModelPort	34.85
DJUI	BasicPort	44.76	BasicPort	44.76
	Non_MP	44.19	Non_MP	44.19
	ModelPort	35.05	ModelPort	37.24
TRCI	BasicPort	44.76	BasicPort	44.76
	Non_MP	31.04	Non_MP	31.04
	ModelPort	27.40	ModelPort	28.98
RI	BasicPort	44.76	BasicPort	44.76
	Non_MP	33.53	Non_MP	33.53
	ModelPort	28.36	ModelPort	29.92
6 Individual Commodity Futures Portfolio	BasicPort	44.76	BasicPort	44.76
	Non_MP	44.78	Non_MP	44.78
	ModelPort	48.22	ModelPort	43.64



〈Figure 1〉 Cumulative Yield Curve of Model Portfolio for Goldman Sachs Commodity Index and 6 Individual Commodity Futures Portfolio

때, 성과가 보다 좋지 않은 것을 확인할 수 있다. 다만 6개 개별선물 동일가중 가격지수를 예측하여 이를 적용한 포트폴리오의 성과는 전통

적 포트폴리오보다 좀 더 나은 것을 확인할 수 있다. 이로써 기계학습을 통한 상품선물 포트폴리오 구성에서는 시장에서 거래가능 상품선물

지수에 투자하는 것보다 개별상품선물 지수에 투자하는 것이 보다 높은 성과를 이룰 수 있음을 알 수 있다.

4.2 6개의 개별 상품선물

4.1절에는 기계학습을 통한 상품선물 포트폴리오 구성 시 개별상품선물 지수를 이용하는 것이 바람직함을 밝혔다. 4.2절에서는 어떤 종류의 상품선물의 모델링 성과가 좋은 지 파악하기 위해 각각의 개별상품선물의 월 수익률을 예측하여 이를 이용하여 개별상품선물 포트폴리오를 구성하였을 때의 성과를 분석하였다. 모델링 결과 각각의 예측률은 <Table 5>와 같다.

<Table 5>를 통해 각 개별선물의 월 수익률 등락 예측률은 4.1절의 상품지수 예측률보다 전반적으로 높음을 알 수 있다. 즉 상품선물을 이용한 포트폴리오 구성에서 SVM 모델의 성능을 발휘하기 위해서는 개별상품선물을 이용하는 것이 합리적이라는 4.1절의 결과에 또 하나의 타당성을 더해준다.

한편 <Table 6>에는 각 개별상품선물에 따른 포트폴리오별 누적수익률 값을 커널 함수별로 정리하였다. <Table 6>의 누적수익률 수치는 상품선물 중에서 옥수수, 밀 등의 곡물 선물, 금, 은

등의 귀금속 선물, WTI, 천연가스 등의 에너지 선물 순서로 모델 적용 포트폴리오의 성과가 우수함을 설명한다. 또한 4.1절과 마찬가지로 가시적인 분석결과 파악을 위해 각 상품섹터를 대표하는 WTI, 옥수수, 은 선물의 각 포트폴리오의 누적수익률 곡선을 <Figure 2>에 나타내었다. 커널 함수는 해당 상품에 대하여 보다 좋은 예측력을 보였던 Radial Basis 커널 함수에 대한 누적수익률 곡선만을 나타냈다.

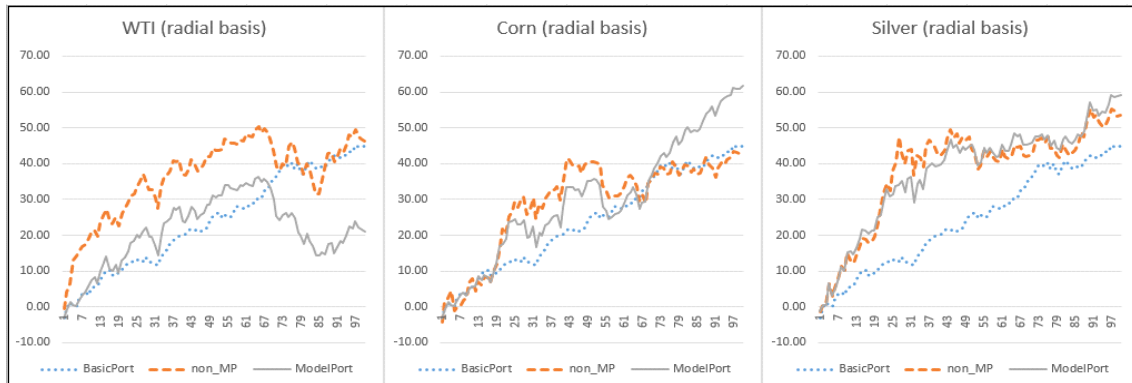
<Figure 2>의 결과를 분석해보면 에너지 선물인 WTI의 경우 모델에 의해 구성된 포트폴리오의 성과가 전통적 포트폴리오의 성과에 하회하는 것을 알 수 있다. 실제로 에너지는 경우 경제변수보다 산유국의 정치적 영향을 많이 받는 상품이기 때문에 단순히 경제변수만으로 예측하기에 한계가 있다고 판단이 되며 실제로 결과로써 이 점이 증명된다. 에너지 상품을 대표하는 석유의 경우 1970년대 탄생한 OPEC과 비OPEC의 갈등과 중동의 정치적 상황 및 최근 미국 셰일가스(Shale Gas) 혁명 이후 원유수출 재개 등 에너지 가격은 세계경기의 변동보다는 산유국의 정책적 요소가 더욱 강하게 미치고 이에 따라 경기가 후행 하여 반응하기 때문에 경제변수를 통하여 예측하는 모델에서는 한계가 있다는 판단이다. 반

(Table 5) Accuracy of 6 Individual Commodity Futures

Commodity Futures	Accuracy(Linear)	Accuracy(Radial)
WTI Crude Oil Futures	51%	51%
Natural Gas Futures	51%	53%
Corn Futures	56%	62%
Wheat Futures	58%	52%
Gold Futures	58%	53%
Silver Futures	49%	52%

〈Table 6〉 Cumulative Yield of 6 Individual Commodity Futures Portfolio

Commodity Futures	Linear		Radial Basis	
	Portfolio	Cumulative Yield	Portfolio	Cumulative Yield
WTI Crude Oil Futures	BasicPort	44.76	BasicPort	44.76
	Non_MP	46.39	Non_MP	46.39
	ModelPort	42.00	ModelPort	20.97
Natural Gas Futures	BasicPort	44.76	BasicPort	44.76
	Non_MP	41.35	Non_MP	46.39
	ModelPort	40.73	ModelPort	46.28
Corn Futures	BasicPort	44.76	BasicPort	44.76
	Non_MP	43.63	Non_MP	43.63
	ModelPort	52.55	ModelPort	61.90
Wheat Futures	BasicPort	44.76	BasicPort	44.76
	Non_MP	38.93	Non_MP	38.93
	ModelPort	54.50	ModelPort	46.39
Gold Futures	BasicPort	44.76	BasicPort	44.76
	Non_MP	44.80	Non_MP	44.80
	ModelPort	51.20	ModelPort	56.31
Silver Futures	BasicPort	44.76	BasicPort	44.76
	Non_MP	53.56	Non_MP	53.56
	ModelPort	53.56	ModelPort	59.10



〈Figure 2〉 Cumulative Yield Curve of Model Portfolio for WTI, Corn, Silver Commodity Futures

면 금속선물은 선물 포함 포트폴리오 및 모델에 의해 구성된 포트폴리오의 성과가 전통적 포트폴리오의 성과보다 미미하게 좋은 것을 확인할

수 있으며 곡물 선물의 경우는 모델에 의해 구성된 포트폴리오의 성과가 전통적 포트폴리오에 비해 상당히 좋은 것을 확인할 수 있다.

따라서 4.2절의 연구 결과는 경제변수만을 Input Data로 하여 SVM 모델링을 하는 경우에는 경제변수 외에 여러 가지 요인에 의해 영향을 받는 에너지 상품을 Output 변수에서 제외하는 것이 바람직하다는 것을 보여준다.

4.3 매 기간별로 개별 상품선물의 등락을 예측하여 상승예측 상품선물만을 편입

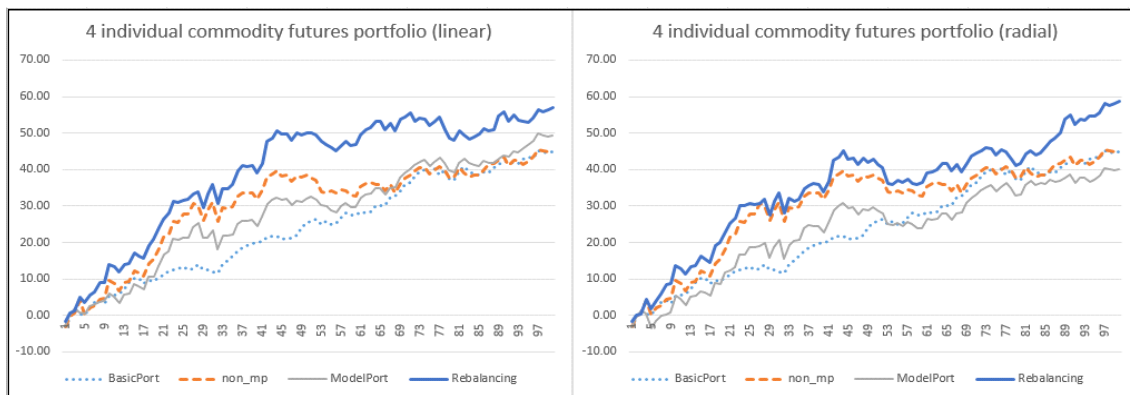
4.3절에서 소개할 포트폴리오 편입 상품선물의 형태는 모델에 의해 매 기간 선물 구성자산이 바뀌는 구조이다. 아이디어는 Input Data를 이용하여 개별 상품선물의 수익률을 예측했던 4.2절

의 결과를 참고하여 에너지 상품선물을 제외한 4개의 상품 선물 중 각 기간별로 'Up'으로 예측한 개별상품 선물들의 동일가중 지수를 구한 후 이를 전통적 포트폴리오에 추가하여 성과를 분석하는 방법이다. 매 기간마다 서로 다른 상품선물들로 선물 자체의 포트폴리오를 구성하기 때문에 이를 리밸런싱 포트폴리오(Rebalancing Portfolio, Rebalancing)라고 칭한다.

<Table 7>에는 커널 함수 별 포트폴리오 누적 수익률 값을 정리하였으며, 누적수익률 곡선을 <Figure 3>에 나타냈다. 한편 <Table 7>에는 기간별 누적수익률과 기간별 수익률의 변동성을

<Table 7> Cumulative Yield and Sharpe Ratios of 4 Individual Commodity Futures Portfolio including Rebalancing Portfolio

Linear				Radial Basis			
4 Individual Commodity Futures	Portfolio	Cumulative Yield	Sharpe Ratio	4 Individual Commodity Futures	Portfolio	Cumulative Yield	Sharpe Ratio
	BasicPort	44.76	0.29		BasicPort	44.76	0.29
	Non_MP	45.23	0.17		Non_MP	45.23	0.17
	ModelPort	49.47	0.23		ModelPort	40.13	0.16
	Rebalancing	57.05	0.23		Rebalancing	58.61	0.24



<Figure 3> Cumulative Yield Curve and Sharpe Ratios of 4 Individual Commodity Futures Portfolio Including Rebalancing Portfolio

고려한 샤프 비율(Sharpe Ratio, SR) 값을 별도의 열에 추가하였다. 샤프 비율 값은 무위험수익률에 대한 초과수익률의 평균을 수익률의 변동성으로 나눈 값으로써 수익성을 객관적으로 나타내는 지표이다. 4.3절에서의 실험 기간의 전통적 포트폴리오의 누적수익률은 커널의 종류와 관계없이 가장 낮지만 해당 기간의 전통적 포트폴리오의 수익률 변동성이 다른 포트폴리오보다 상대적으로 낮기에 샤프 비율이 0.29로 가장 높은 것을 알 수 있다. 하지만 선물이 포함된 포트폴리오 중에서는 누적수익률 및 샤프 비율에서 리밸런싱 포트폴리오의 성과가 가장 우수함을 확인할 수 있다.

4.4 다른 훈련기간 및 실험기간에 대하여 4.3절의 실험을 재구성

앞선 4.1절부터 4.3절까지의 분석은 총 데이터 기간에 대하여 앞부분은 훈련기간으로 뒷부분은 실험기간으로 나누었다. 즉 과거의 데이터로 훈련하여 최근에 데이터로 검증을 실시하였다. 하지만 Output에 대한 Input Data의 설명력, 모델의 타당성을 검증하기 위해서는 총 데이터 중 임의로 훈련기간과 실험기간을 나누어서 모델링 후 결과분석을 하더라도 4.3절과 유사한 분석결과

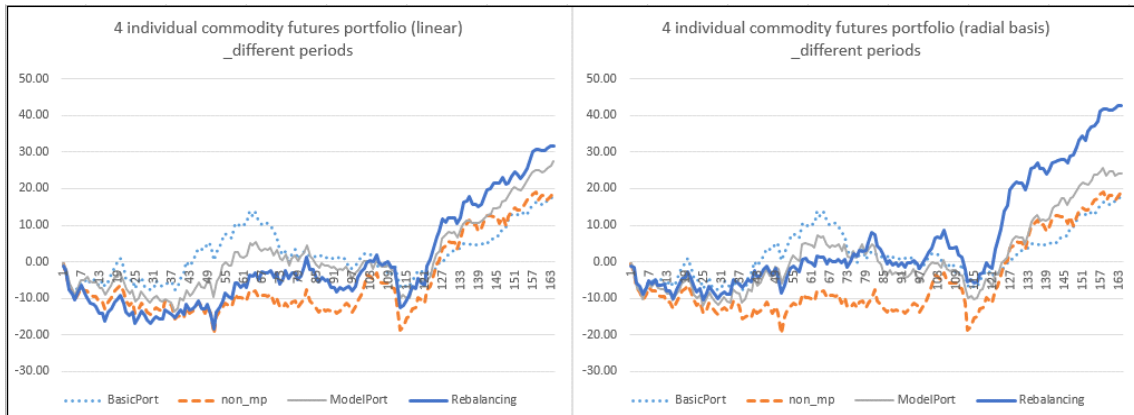
가 나와야 한다고 판단하였고, 총 데이터 기간 중 홀수 년도에 해당하는 기간의 데이터를 훈련 데이터로, 짝수 년도에 해당하는 기간의 데이터를 실험데이터로 분류하여 같은 실험을 진행하였다. <Table 8>에는 커널 함수 별 포트폴리오 누적수익률 값을 정리하였으며, 누적수익률 곡선을 <Figure 4>에 나타냈다.

4.3절 및 4.4절은 리밸런싱 포트폴리오의 성과가 앞에서 소개한 벤치마크나 Non-model Portfolio나 Model Portfolio의 성과보다 더 우월함을 보여준다. 또한 4.4절의 실험 기간에서는 전통적 포트폴리오의 수익성이 미미하였기에 4.3절에 대비하여 샤프 비율에서도 리밸런싱 포트폴리오의 성과가 가장 우수함을 알 수 있다.

4장의 분석 결과를 비추어 볼 때 본 연구에서는 다음과 같은 결론을 내린다. 포트폴리오 성과 향상을 위해 기존에 널리 알려진 전통적 포트폴리오에 SVM 모델을 기반하여 상품 선물을 편입하고자 할 때에 투자대상은 상품지수에 투자하는 것이 아닌 개별 상품선물을 선정하여 자체적 상품선물 포트폴리오를 구성하고 그 방법으로는 기간마다 강세가 예측되는 개별 선물만을 골라서 포트폴리오를 재구성하는 것이 바람직하다는 것이다.

<Table 8> Cumulative Yield and Sharpe Ratios of 4 Individual Commodity Futures Portfolio including Rebalancing Portfolio for Different Training Periods and Test Periods

Linear				Radial Basis			
	Portfolio	Cumulative Yield	Sharpe Ratio		Portfolio	Cumulative Yield	Sharpe Ratio
4 Individual Commodity Futures	BasicPort	17.41	0.06	4 Individual Commodity Futures	BasicPort	17.41	0.06
	Non_MP	19.21	0.06		Non_MP	19.21	0.06
	ModelPort	27.48	0.10		ModelPort	24.18	0.08
	Rebalancing	31.57	0.10		Rebalancing	42.58	0.14



(Figure 4) Cumulative Yield Curve and Sharpe Ratios of 4 Individual Commodity Futures Portfolio Including Rebalancing Portfolio for Different Training Periods and Test periods

5. 결론 및 향후 연구

전통자산의 투자효율성을 증대 시키고 포트폴리오의 안정성과 수익성을 향상 시키기 위한 상품자산에 대한 분산투자의 가치는 투자자들에게 보편적으로 인정받는 사항이다. 본 연구는 이러한 상품자산에 대한 실질적인 투자전략을 연구해 보았다. 본 연구에서는 상품자산 투자 예측을 위해 머신 러닝 기법을 사용하였기에 기존의 연구들과 차별성을 가지고 있다.

본 연구를 통하여 발견한 점은 시장에서 널리 알려진 상품지수에 투자하는 것보다 개별상품선물 동일비중으로 구성된 포트폴리오를 이용하는 것이 좋은 성과를 나타냈으며, 더 나아가 개별 상품선물 각각을 예측한 모델을 기반으로 구성된 포트폴리오가 더 우수한 성과를 나타냈다. 개별 상품선물의 예측성과를 살펴보았을 때 옥수수과 밀과 같은 농산품이 상대적으로 높은 성과를 나타냈으며 경제적 변수보다 산유국의 정책적 변수에 많은 영향을 받는 에너지 상품은 상대

적으로 저조한 성과를 나타냈다. 따라서 에너지 상품을 제외한 상품선물 각각 예측모델을 통하여 구성된 포트폴리오로 상품투자를 행하는 경우가 가장 우수한 성과를 나타냈다. 본 연구에서는 실제 투자가 가능한 상품지수와 상품선물에 대한 매매전략을 연구했기 때문에 실제 자산배분 전략에 응용할 수 있다는 측면에서도 연구의 의미가 있다.

본 연구는 상품지수와 개별상품을 중심으로 연구되었기에 상품ETF나 다른 상품자산을 이용한 상품들에 대한 검토가 부족한 점이 있다. 또한 국내 투자자들이 해외CTA들에 대한 투자가 활발해 진다면 해외CTA들에 대한 자산배분의 연구도 이 연구를 바탕으로 추가적으로 이루어질 수 있을 것이다. 본 연구는 자산배분의 입장에서 월별데이터를 사용하고 다소 긴 기간을 분석했으나 본 연구를 바탕으로 보다 짧은 기간의 각 상품자산의 특성을 파악하여 투자전략을 연구한다면 보다 실질적인 투자모델을 수립할 수 있을 것이다.

더 나아가 경기 상승 및 하락 국면, 금리 상승 및 하락 국면, 주식시장 상승 및 하락 국면 등 국면 별 시장상황에 따른 분석 등이 추가된 자산배분 투자모델을 설계한다면 자산배분에 대한 지능형 서비스(Intelligent Service)도 금융시장에서 만날 수 있게 될 것이다.

참고문헌(References)

- Bhardwaj, G., G. Gorton, and K. G. Rouwenhorst, "Fooling Some of the People All of the Time: The Inefficient Performance and Persistence of Commodity Trading Advisors," *The Review of Financial Studies*, Vol.27, No.11(2014), 3099~3132.
- Bhardwaj, G., G. Gorton, and K. G. Rouwenhorst, *Facts and Fantasies About Commodity Futures Ten Years Later*, 2015. Available at <http://faculty.som.yale.edu/garygorton/documents/FactsandFantasiesaboutCommodity.pdf> (Accessed 2 October, 2017).
- Boser, B. E., I. M. Guyon, and V. N. Vapnik, *A Training Algorithm for Optimal Margin Classifiers*, COLT '92 Proceedings of the fifth annual workshop on Computational learning theory, 1992. Available at <http://dl.acm.org/citation.cfm?doid=130385.130401> (Accessed 2 October, 2017).
- Cha, E. J. and T. H. Hong, "Stock Index Prediction Using SVM and News Sentimental Analysis," *The Korea Society of Management Information Systems*, Vol.2016, No.6(2016), 64~75.
- Choi, H. S., S. W. Kim, and S. C. Park, "Analysis of Trading Performance on Intelligent Trading System for Directional Trading," *Journal of Intelligence and Information Systems*, Vol.17, No.3(2011), 187~201.
- CME Group, *Managed Futures: Portfolio Diversification Opportunities*, CME Group, 2012.
- Flach, P., *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, Cambridge University Press, 2012.
- Gorton, G. and K. G. Rouwenhorst, "Facts and Fantasies about Commodity Futures," *Financial Analysts Journal*, Vol.62, No.2(2006), 47~68.
- Heo, J. Y. and J. Y. Yang, "SVM based Stock Price Forecasting Using Financial Statements," *KIISE Transactions on Computing Practices*, Vol.21, No.3(2015), 167~172.
- Kim, S. W. and H. C. Ahn, "Development of an Intelligent Trading System using Support Vector Machines and Genetic Algorithms," *Journal of Intelligence and Information Systems*, Vol.16, No.1(2010), 71~92.
- Kim, Y. S., N. G. Kim, and S. R. Jeong, "Stock-Index Invest Model Using News Big Data Opinion Mining," *Journal of Intelligence and Information Systems*, Vol.18, No.2(2012), 143~156.
- Lantz, B., *Machine Learning with R*. Packt, Birmingham, UK, 2013.
- Li, X., D. Lord, Y. Zhang, and Y. Xie, "Predicting motor vehicle crashes using Support Vector Machine models," *Accident Analysis & Prevention*, Vol.40, No.4(2008), 1611~1618.
- Lintner, V. J., *The Potential Role of Managed Commodity-financial Futures Accounts*

- (and/or Funds) in Portfolios of Stocks and Bonds, Harvard university, 1983.
- Oh, I. S., *Pattern Recognition*, Kyobobook, 2008.
- Park, K. H., T. Hou, and H. J. Shin, "Oil Price Forecasting Based on Machine Learning Techniques," *Journal of the Korean Institute of Industrial Engineers*, Vol.37, No.1(2011), 64~73.
- Park, S. H. and B. Hansen, "Prediction of Protein-Protein Interaction Sites Based on 3D Surface Patches Using SVM," *Korea Information Processing Society*, Vol.19, No.1(2012), 21~28.
- Ra, Y. S., H. S. Choi, and S.W. Kim, "VKOSPI Forecasting and Option Trading Application Using SVM," *Journal of Intelligence and Information Systems*, Vol.22, No.4(2016), 177~192.
- Kim, S. W. and H. S. Choi, "Estimation of GARCH Models and Performance Analysis of Volatility Trading System Using Support Vector Regression," *Journal of Intelligence and Information Systems*, Vol.23, No.2(2017), 123~138.
- Ramakrishnan, S., S. Butt, M. A. Chohan, and H. Ahmad, *Forecasting Malaysian exchange rate using machine learning techniques based on commodities prices*, 2017 International Conference on Research and Innovation in Information Systems (ICRIIS), 2017.
- Tay, F. E. H. and L. J. Cao, "Modified support vector machines in financial time series forecasting," *Neurocomputing*, Vol.48, No.1-4(2002), 847~861.
- Xie W., L. Yu, S. Xu, and S. Wang, *A New Method for Crude Oil Price Forecasting Based on Support Vector Machines*, ICCS 2006: Computational Science, 2006.

Abstract

A Study on Commodity Asset Investment Model Based on Machine Learning Technique

Song, Jin Ho* · Choi, Heung Sik** · Kim, Sun Woong*

Services using artificial intelligence have begun to emerge in daily life. Artificial intelligence is applied to products in consumer electronics and communications such as artificial intelligence refrigerators and speakers. In the financial sector, using Kensho's artificial intelligence technology, the process of the stock trading system in Goldman Sachs was improved. For example, two stock traders could handle the work of 600 stock traders and the analytical work for 15 people for 4weeks could be processed in 5 minutes. Especially, big data analysis through machine learning among artificial intelligence fields is actively applied throughout the financial industry.

The stock market analysis and investment modeling through machine learning theory are also actively studied. The limits of linearity problem existing in financial time series studies are overcome by using machine learning theory such as artificial intelligence prediction model. The study of quantitative financial data based on the past stock market-related numerical data is widely performed using artificial intelligence to forecast future movements of stock price or indices. Various other studies have been conducted to predict the future direction of the market or the stock price of companies by learning based on a large amount of text data such as various news and comments related to the stock market.

Investing on commodity asset, one of alternative assets, is usually used for enhancing the stability and safety of traditional stock and bond asset portfolio. There are relatively few researches on the investment model about commodity asset than mainstream assets like equity and bond. Recently machine learning techniques are widely applied on financial world, especially on stock and bond investment model and it makes better trading model on this field and makes the change on the whole financial area.

In this study we made investment model using Support Vector Machine among the machine learning models. There are some researches on commodity asset focusing on the price prediction of the specific

* Graduate School of Business IT, Kookmin University
** Corresponding Author: Choi, Heung Sik
Graduate School of Business IT, Kookmin University
77 Jeongneung-ro, Seongbuk-gu, Seoul 02707, Korea
Tel: +82-2-910-4567, Fax: +82-2-910-4017, E-mail: hschoi@kookmin.ac.kr

commodity but it is hard to find the researches about investment model of commodity as asset allocation using machine learning model. We propose a method of forecasting four major commodity indices, portfolio made of commodity futures, and individual commodity futures, using SVM model. The four major commodity indices are Goldman Sachs Commodity Index(GSCI), Dow Jones UBS Commodity Index(DJUI), Thomson Reuters/Core Commodity CRB Index(TRCI), and Rogers International Commodity Index(RI). We selected each two individual futures among three sectors as energy, agriculture, and metals that are actively traded on CME market and have enough liquidity. They are Crude Oil, Natural Gas, Corn, Wheat, Gold and Silver Futures. We made the equally weighted portfolio with six commodity futures for comparing with other commodity indices.

We set the 19 macroeconomic indicators including stock market indices, exports & imports trade data, labor market data, and composite leading indicators as the input data of the model because commodity asset is very closely related with the macroeconomic activities. They are 14 US economic indicators, two Chinese economic indicators and two Korean economic indicators. Data period is from January 1990 to May 2017. We set the former 195 monthly data as training data and the latter 125 monthly data as test data.

In this study, we verified that the performance of the equally weighted commodity futures portfolio rebalanced by the SVM model is better than that of other commodity indices. The prediction accuracy of the model for the commodity indices does not exceed 50% regardless of the SVM kernel function. On the other hand, the prediction accuracy of equally weighted commodity futures portfolio is 53%. The prediction accuracy of the individual commodity futures model is better than that of commodity indices model especially in agriculture and metal sectors. The individual commodity futures portfolio excluding the energy sector has outperformed the three sectors covered by individual commodity futures portfolio. In order to verify the validity of the model, it is judged that the analysis results should be similar despite variations in data period. So we also examined the odd numbered year data as training data and the even numbered year data as test data and we confirmed that the analysis results are similar. As a result, when we allocate commodity assets to traditional portfolio composed of stock, bond, and cash, we can get more effective investment performance not by investing commodity indices but by investing commodity futures. Especially we can get better performance by rebalanced commodity futures portfolio designed by SVM model.

Key Words : Commodity Asset, Machine Learning, Support Vector Machine, Investment Model

Received : October 11, 2017 Revised : November 22, 2017 Accepted : December 1, 2017

Publication Type : Regular Paper Corresponding Author : Choi, Heung Sik

저 자 소개



송진호

연세대학교 경영학과에서 학사 및 석사학위를 취득하였고, 현재 국민대학교 비즈니스IT 전문대학원에서 트레이딩시스템 전공 박사과정에 재학 중이다. 관심분야로 자산배분, 대체투자, 헤지펀드, 기계학습, 금융공학 등이 있다.



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, 트레이딩계량 분석, 옵션 변동성매매 등이다.



김선웅

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