

## 비즈니스 분석학적 접근을 통한 업종별 ETF 등락 예측모형\*

강 동 철\*\* · 김 원 경\*\*\* · 오 정 석\*\*\*\*

《目 次》

- |          |         |
|----------|---------|
| I. 서 론   | III. 분석 |
| II. 연구방법 | IV. 결 론 |

### I. 서 론

오늘날 수많은 정보가 다양한 매체를 통해 수집되고 쏟아져 나온다. 이러한 추세는 오늘날 인터넷뿐만 아니라 모바일, 사물인터넷 등의 급속한 발전으로 데이터의 축적 속도가 더욱 높아지고 있다. 각 정보들은 그 자체로서의 의미보다는 가공, 분석 과정을 통해 실제 비즈니스에 도움이 되는 정보로 변화한다. 비즈니스 분석학은 이러한 과정에서 중심적인 역할을 한다. 경영과학기법을 활용해 빅데이터를 기반으로 기업가치를 증대하기 위해 사용되는 방법으로서, 분야에서 비즈니스 분석학의 필요성이 대두되면서 다양한 분야에서 이를 적용한 연구, 비즈니스 사례에 대한 연구가 활발히 이루어지고 있다.

주식시장의 흐름을 예측하고자 하는 다양한 노력들이 있었다. 김삼용 등(2009)은 일반 자기회귀 이분산 모형을 이용하여 KOSPI의 수익률에 대한 최적의 모형을 구축하고자 하였다. 육근철 등(1999)은 인공 신경망을 이용하여 KOSPI 200지수를 예측하기 위해 선행적으로 산업별 주가 지수를 예측하는 상향식 예측 모형을 수립하였다. 이형용(2008)은 KOSPI 지수 등락 예측을 위해 구축된 로지스틱 회귀모형, 인공신경망, SVM모형의 결과들을 제안모형을 이용해 결합하여 새로운 모형을 구축하였다. 이희준(2004)은 Kodex200 지수에 대해 영향을 미치는 변수를 선정하

\* 본 연구는 서울대학교 경영정보연구소의 연구비 지원에 의해 이루어졌습니다.

\*\* 서울대학교 경영학과 학사과정

\*\*\* 서울대학교 소비자학과 학사과정

\*\*\*\* 서울대학교 경영대학 교수

고 신경망 분석과 로지스틱 회귀분석을 이용하여 주가 수익률을 예측하고 비교 분석하였다.

하지만 이를 예측하는데 있어 각 산업마다 변수들에 의해 영향을 받는 정도와 관련된 변수들이 모두 다르며, 국내뿐만 아니라 외국의 경제적 요인에 영향을 받는다는 한계점이 존재한다. 예를 들어 고희운(2016)은 국제원유의 충격이 국내 산업별 주가지수와 직접적으로 영향을 받고 민감하게 반응함을 BEKK모형을 이용하여 확인하였다. 빙기범 등(2007)은 다양한 해외 요인들이 한국 증시의 수익률과 변동성에 지대한 영향을 미친다는 것과, 중국 경제 및 증시가 가져오는 특성을 분석하였다. 이러한 한계점을 해결하기 위하여 최익선 등(2012)은 국내외 경제지표를 변수로 활용하여 로지스틱 회귀모형과 신경망 모형을 적용하여 산업별 주가지수의 상승 및 하락을 예측하였는데, 해외 경제상황에 큰 영향을 받는 우리나라 주식 시장을 고려하여 국내 경제지표뿐만 아니라 미국, 일본, 중국, 유럽의 주요 경제지표를 예측변수로 사용하였다. 본 연구에서도 모든 산업을 통합하여 예측하지 않고 산업별 주가지수의 흐름을 살펴보고자 한다. 이를 위해 본 연구는 실제로 거래로 이어질 수 있는 산업별 상장지수펀드(ETF)의 등락을 예측하고자 한다. ETF란 지수의 움직임과 동일한 수익률을 내도록 만들어진 지수연동형펀드의 일종이다. 업종별 ETF는 특정 업종에 연동하는 것으로서 이를 통해 업종에 대한 전반적인 흐름을 파악할 수 있다. 또한 변수들에 있어서도 국내 경제지표뿐만 아니라 해외의 경제지표 또한 예측변수로 사용함으로써 예측의 정확도를 높이고자 한다. 이를 통해 모든 산업을 통합하여 예측하지 않고 업종별 상장지수펀드를 예측하고자 한다. 개별 주식에 대한 예측을 하기보다는 각 개별 주식들이 모여 만들어진 지수 연동형 펀드를 예측함으로써 예측오차를 보다 줄일 수 있다. 또한 변수들에 있어서도 국내 경제지표뿐만 아니라 해외의 경제지표 또한 예측변수로 사용함으로써 예측의 정확도를 높일 수 있다.

본 연구에서는 분석을 위해 각 ETF가 발행한 시점부터 2015년까지의 월별자료를 목표변수로 사용하였다. 총 38개의 국내외 경제변수들을 단계적 변수선택법에 의해 추출하여 로지스틱 회귀모형을 사용하여 ETF의 전월대비 상승, 하락을 예측하였다. 이에 사용된 통계 프로그램은 SPSS ver21을 사용하였다.

## II. 연구방법

### 2.1 로지스틱 회귀분석

로지스틱 회귀분석은 독립 변수의 선형 결합으로 종속 변수를 설명한다는 점에서 선형 회귀 분석과 유사하다. 그러나 로지스틱 회귀 분석은 선형회귀 분석과는 다르게 종속 변수가 범주형 데이터를 대상으로 하며 입력 데이터가 주어졌을 때 해당 데이터의 결과가 특정 분류로 나뉘기 때문에

일종의 분류 기법으로 볼 수 있다. 로지스틱 회귀 분석의 종류는 이항형 혹은 다항형이 있는데, 이 연구에서는 이항형 로지스틱 회귀(binomial logistic regression)분석을 사용한다. 이항형 로지스틱 회귀 분석은 종속 변수의 결과가 (1,0) 또는 (성공,실패)와 같이 2개의 카테고리가 존재하는 것을 의미한다. 이 연구에서는 ETF가 상승했을 경우를 1, 하락했을 경우를 0으로 값을 부여하여 분석을 진행하였다.

데이터 마이닝 기법중 하나인 로지스틱 회귀 분석은 과적합화(overfitting)의 위험성을 갖고 있다. 즉, 학습용 데이터를 통해 얻어진 모형이 그 데이터에 너무 가깝게 적합화되면 이 모형은 데이터가 갖는 구조적 특성뿐만 아니라 우연적 특수성을 모두 반영하게 되어 과적합화 현상이 발생하게 된다. 이러한 과적합화 현상은 변수선택법을 사용함으로써 해결할 수 있는데(Stephan Dreiseitl 등(2002)), 그 중 본 연구에서는 가능도비를 이용한 전진단계 선택법 혹은 후진단계 선택법을 사용함으로써 해결하려 하였다. 전진단계 선택법은 회귀식에 포함되는 변수 중 가장 의미있는 변수부터 단계별로 유의성을 검증하는 방법으로, 현재 모형의 최대가능도 추정량에 기초하여 각 변수들의 유의성을 검증한 후 모형에 포함시킬 변수를 하나씩 선택하는 방법이다. 후진단계 선택법은 전진단계 선택법과 정반대로 모든 변수를 포함한 회귀식으로부터 단계별로 변수를 제거해나가는 방법이다. 이 연구는 각각의 ETF에 대해 두 개의 방법을 모두 사용하여 분석한 뒤, 그 중에서 설명력이 큰 방법을 이용하여 모형으로 제시하였다.

## 2.2 연구모형

### 2.2.1 종속변수의 선택

종속변수로는 우리나라 산업을 대표할 수 있는 10개의 산업에 해당하는 ETF를 선택하였으며, 2006년부터 2016년까지 11년 동안의 월별자료를 사용할 예정이다. 자료중에서 일부분은 모형화를 위한 자료로 사용하고, 나머지는 예측용 자료로 사용하였다. 다음은 선택한 종속변수와 그에 대한 간략한 소개이다.

#### 1) Tiger 반도체

유가증권시장 및 코스닥시장에 상장된 종목중 전기.전자 산업에 속한 반도체칩 제조기업과 다른 산업에 속한 종목 중 반도체 부품, 장비제조 관련 20여개의 기업으로 구성된 KRX Semicon 지수의 변동률과 유사하도록 구성된 ETF(대표종목 : SK하이닉스, 원익 IPS, 동부하이텍, AP시스템, 서울반도체)

## 2) Kodex 은행

은행업종의 주가흐름을 반영하는 KRX Banks 지수의 변동률과 유사하도록 구성된, 국내 은행 산업을 대표하는 10개의 기업으로 구성된 ETF(대표종목 : 하나금융지주, KB금융, 신한지주, 우리은행)

## 3) Kodex 증권

국내 종합건설업, 건축기술 및 엔지니어링 서비스 업종을 대표하는 10개의 기업으로 구성된 KRX Constructions 지수의 변동률과 유사하도록 구성된 ETF(대표종목 : 현대산업, 현대건설, 대림산업, KCC)

## 4) Kodex 철강

국내 철광업, 비철금속광업, 제 1차 철강산업, 1차 비철금속산업, 금속주조업종을 대표하는 KRX Steels 지수의 변동률과 유사하도록 구성된 ETF(대표종목 : POSCO, 현대제철, 고려아연, 동국제강)

〈표 1〉 독립변수

| 번호  | 변수 이름          | 번호  | 변수 이름                               |
|-----|----------------|-----|-------------------------------------|
| x1  | 한국-수출물가지수      | x20 | 소비자기대지수                             |
| x2  | 한국-설비물가지수      | x21 | 건설수주액                               |
| x3  | 미국-필라델피아 연준 지수 | x22 | 금융기관유동성                             |
| x4  | 미국-시카고 PMI 지수  | x23 | 산업생산지수                              |
| x5  | 미국-소비자 신뢰지수    | x24 | 제조업 가동율 지수                          |
| x6  | 일본-산업생산지수      | x25 | 건설 기성액                              |
| x7  | 중국-소매판매지수      | x26 | 서비스업 생산지수                           |
| x8  | 중국-산업생산지수      | x27 | 생산 자체품 재고지수                         |
| x9  | 미국주가지수-S&P500  | x28 | 기계수주액                               |
| x10 | 일본주가지수-니케이225  | x29 | 자본재수입액                              |
| x11 | 중국주가지수-상해종합    | x30 | 장단기금리차                              |
| x12 | 원화             | x31 | 순상품교역조건                             |
| x13 | 위안화            | x32 | 구인구직비율                              |
| x14 | 유로             | x33 | 경기선행지수 OECD(Amplitude Adjusted)     |
| x15 | 엔화             | x34 | 경기선행지수 OECD(Trend Restored)         |
| x16 | 경기선행지수         | x35 | 경기선행지수 OECD(Europe, Trend Restored) |
| x17 | 경기동행지수         | x36 | 금                                   |
| x18 | 경기동행지수 순환변동치   | x37 | WTI                                 |
| x19 | 재고순환지표         | x38 | 브랜트유                                |

### 2.2.2 독립변수의 설정

앞서 진행된 선행 연구를 바탕으로 국내경제지표와 해외경제지표 40 여개의 변수를 선정하도록 한다. 그리고 로지스틱 분석을 위해 40 여개의 변수를 모두 사용하기에는 자료수에 대비하여 너무 많기 때문에 앞서 2.1에서 설명한 바와 같이 SPSS의 로지스틱 회귀분석 기능에 있는 전진 혹은 후진 로지스틱 회귀 방법을 활용하여 유의한 변수를 추출하여 로지스틱 회귀분석을 실시할 것이다. <표 1>은 선행자료에서 선택한 38개의 입력변수이다.

## III. 분석

### 3.1 분석 방법

2006년부터 2016년까지의 자료들을 모형화를 위한 자료와 예측용 자료로 나누어서 분석을 실시하였다. 모형화의 경우에는 반도체는 2006년 9월~2014년 12월, 은행은 2006년 9월~2014년 12월, 증권은 2008년 8월~2014년 12월, 철강은 2010년 1월~2014년 12월의 자료를 사용하였다. 예측용은 모두 2015년의 1년간의 자료를 사용하였다.

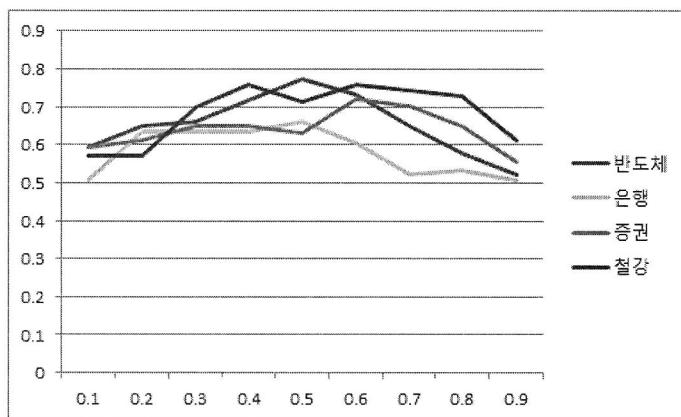
모형화를 위한 절단값 설정을 위하여 절단값에 따른 정확도, 민감도, 특이도에 대한 분석이 필요하다. 정확도는 상승과 하락을, 민감도는 상승을, 특이도는 하락을 정확히 구별했는가에 대한 척도로 이 세 가지의 값이 높을 때 최적화된 절단값을 구할 수 있다(Balkishan Sharma(2014)). 다음의 <표 2>는 반도체 산업의 절단값에 따른 정확도, 특이도, 민감도의 변화 추이이다.

<표 2> 각 산업의 절단값에 따른 정확도, 민감도, 특이도의 변화 추이

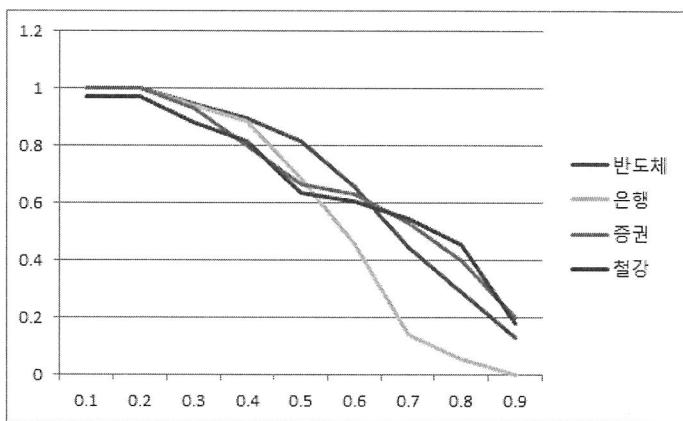
|     | 절단값 | 정확도   | 민감도   | 특이도   |
|-----|-----|-------|-------|-------|
| 반도체 | 0.1 | 0.592 | 1     | 0.121 |
|     | 0.2 | 0.648 | 1     | 0.242 |
|     | 0.3 | 0.662 | 0.947 | 0.333 |
|     | 0.4 | 0.718 | 0.895 | 0.515 |
|     | 0.5 | 0.775 | 0.816 | 0.727 |
|     | 0.6 | 0.732 | 0.658 | 0.818 |
|     | 0.7 | 0.648 | 0.447 | 0.879 |
|     | 0.8 | 0.577 | 0.289 | 0.909 |
|     | 0.9 | 0.521 | 0.132 | 0.970 |

〈표 2〉 각 산업의 절단값에 따른 정확도, 민감도, 특이도의 변화 추이 (계속)

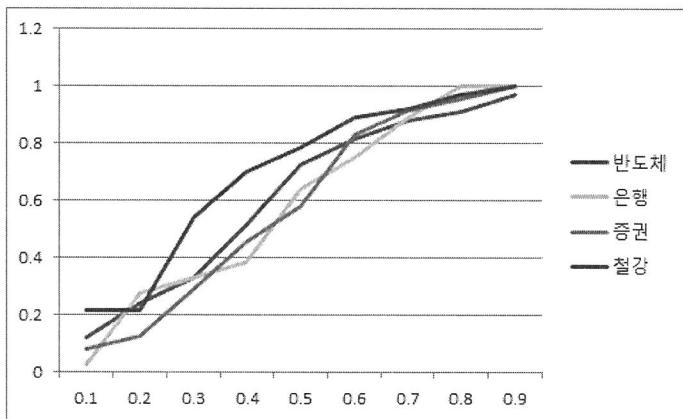
|             | 절단값 | 정확도   | 민감도   | 특이도   |
|-------------|-----|-------|-------|-------|
| 은행          | 0.1 | 0.507 | 1     | 0.028 |
|             | 0.2 | 0.634 | 1     | 0.278 |
|             | 0.3 | 0.634 | 0.943 | 0.333 |
|             | 0.4 | 0.634 | 0.886 | 0.386 |
|             | 0.5 | 0.662 | 0.686 | 0.639 |
|             | 0.6 | 0.606 | 0.457 | 0.750 |
|             | 0.7 | 0.521 | 0.143 | 0.889 |
|             | 0.8 | 0.535 | 0.057 | 1     |
|             | 0.9 | 0.507 | 0     | 1     |
| 증권          | 0.1 | 0.593 | 1     | 0.083 |
|             | 0.2 | 0.611 | 1     | 0.125 |
|             | 0.3 | 0.648 | 0.933 | 0.292 |
|             | 0.4 | 0.648 | 0.800 | 0.458 |
|             | 0.5 | 0.630 | 0.667 | 0.583 |
|             | 0.6 | 0.722 | 0.633 | 0.833 |
|             | 0.7 | 0.704 | 0.533 | 0.917 |
|             | 0.8 | 0.648 | 0.400 | 0.958 |
|             | 0.9 | 0.556 | 0.200 | 1     |
| 철강<br>0.405 | 0.1 | 0.571 | 0.970 | 0.216 |
|             | 0.3 | 0.700 | 0.879 | 0.541 |
|             | 0.4 | 0.757 | 0.818 | 0.703 |
|             | 0.5 | 0.714 | 0.634 | 0.784 |
|             | 0.6 | 0.757 | 0.606 | 0.892 |
|             | 0.7 | 0.743 | 0.545 | 0.919 |
|             | 0.8 | 0.729 | 0.455 | 0.973 |
|             | 0.9 | 0.614 | 0.182 | 1     |



〈그림 1〉 절단값에 따른 업종별 정확도 분석



〈그림 2〉 절단값에 따른 업종별 민감도 분석



〈그림 3〉 절단값에 따른 업종별 특이도 분석

위와 같은 분석을 네 개 각각의 항목에 적용하여 얻어진 절단값은 반도체와 은행은 0.5, 증권과 철강은 각각 0.4, 0.6이다.

### 3.2 모형화

#### 3.2.1 반도체

로지스틱 회귀 모형 :

$$p = \frac{1}{1 + EXP(-( -1.035 * 경기선행 + 1.466 * 경기동행 - 2.822 * 경기동행순환변동 + 0.266 * 제조업가동율 + 5.057 * 장단기금리차 + 207.719)))}$$

〈표 3〉 반도체 로지스틱 회귀분석 결과분류표

| 관측번호   |             | 예측값                |      |    | 정정 백분율 |
|--------|-------------|--------------------|------|----|--------|
|        |             | tiger반도체 코딩<br>.00 | 1.00 |    |        |
| 32 단계  | tiger반도체 코딩 | .00                | 24   | 9  | 72.7   |
|        |             | 1.00               | 6    | 32 | 84.2   |
| 전체 백분율 |             |                    |      |    | 78.9   |
| 33 단계  | tiger반도체 코딩 | .00                | 24   | 9  | 72.7   |
|        |             | 1.00               | 7    | 31 | 81.6   |
| 전체 백분율 |             |                    |      |    | 77.5   |

a. 잘라낸 값은 .500임

〈표 4〉 반도체 로지스틱 회귀분석 결과 방정식의 변수

|       | B        | S.E.    | Wald   | Df    | 유의수준 | Exp(B)    |
|-------|----------|---------|--------|-------|------|-----------|
| 32 단계 | 미국시카고    | .071    | .071   | 1     | .317 | 1.073     |
|       | 경기선행     | -1.225  | .488   | 6.302 | 1    | .012      |
|       | 경기동행     | 1.650   | .612   | 7.258 | 1    | .007      |
|       | 경기동행순환변동 | -3.395  | 1.121  | 9.177 | 1    | .002      |
|       | 제조업가동율   | .232    | .128   | 3.299 | 1    | .069      |
|       | 장단기금리차   | 5.057   | 1.987  | 6.477 | 1    | .011      |
| 33 단계 | 상수       | 264.808 | 94.941 | 7.780 | 1    | .005      |
|       | 경기선행     | -1.035  | .433   | 5.708 | 1    | .017      |
|       | 경기동행     | 1.466   | .566   | 6.717 | 1    | .010      |
|       | 경기동행순환변동 | -2.822  | .908   | 9.669 | 1    | .002      |
|       | 제조업가동율   | .266    | .125   | 4.532 | 1    | .033      |
|       | 장단기금리차   | 5.057   | 1.948  | 6.738 | 1    | .009      |
| 상수    |          | 207.719 | 70.847 | 8.596 | 1    | .003      |
|       |          |         |        |       |      | 1.626E+90 |

38개의 변수들 중 설명력이 높은 변수를 채택하기 위해 후진단계 선택법으로 변수 채택하였다. 채택된 변수는 경기선행, 경기동행, 경기동행 순환변동, 제조업가동율, 장단기금리차이다. 분석결과, 경기선행지수와 경기동행순환변동은 주가에 부의 영향을, 경기동행과 제조업가동율, 장단기금리차는 주가에 양의 영향을 미친다고 할 수 있다. 즉 경기선행지수가 1 증가할수록 0.355배 정도 주가 하락할 가능성이 생기고 경기동행순환변동이 1증가할수록 0.059배 주가가 하락할 가능성이 생긴다. 또한, 경기동행지수가 1증가할수록 4.331배 주가가 상승할 가능성이 생기고, 제조업가동율이 1증가할수록 1.304배 주가가 상승할 가능성이 생기며, 장단기금리차가 1 증가할수록 157.184

배 주가가 상승할 가능성이 높아진다. 가장 큰 영향을 받는 변수는 장단기 금리차이다. 장단기금리 차는 경기선행지수로서 민간의 경기 회복에 대한 예상을 담은 기대지수이다. 장단기금리차가 커지는 것은 일반가계에서 이자율 상승을 예측하고 경기 호전을 기대한다는 것을 의미하며 이는 위축된 소비나 투자심리를 살아나게 하여 경기회복을 가져오는 효과가 있다. 이러한 장단기금리차가 반도체 ETF주가에 큰 영향을 미치는 것은 반도체 산업이 국내의 경기변수에 큰 영향을 받는 것이라고 할 수 있다.

### 3.2.2 은행

로지스틱 회귀 모형 :

$$p = 1/(1 + \text{EXP}(-(-0.002*\text{상해종합주가지수}-0.912*\text{경기동행순환변동} \\ + 0.095*\text{소비자기대지수}+85.107)))$$

38개의 변수들 중 설명력이 높은 변수를 채택하기 위해 전진단계 선택법으로 변수 채택한 결과, 채택된 변수는 중국주가상해, 경기동행순환변동, 소비자기대지수이다. 상해중국주가와 경기동행순환변동은 은행ETF 주가에 부의 영향을, 소비자기대지수는 주가에 양의 영향을 미친다고 할 수 있다. 즉 중국주가상해가 1증가할 때 은행ETF의 주가가 0.998배 하락할 가능성이 생기고 경기동행순환변동이 1하락할 때 은행 ETF의 주가가 0.402배 하락할 가능성이 생긴다. 반면 소비자기대지수가 1증가하면 은행 ETF의 주가가 1.100배 증가할 가능성이 생긴다고 할 수 있다.

〈표 5〉 은행 로지스틱 회귀분석 결과 분류표

| 관측번호   |        | 예측값    |    | 정정 백분율 |      |
|--------|--------|--------|----|--------|------|
|        |        | 은행상승하락 |    |        |      |
|        |        | 0      | 1  |        |      |
| 1 단계   | 은행상승하락 | 0      | 22 | 14     | 61.1 |
|        |        | 1      | 18 | 17     | 48.6 |
| 전체 백분율 |        |        |    | 54.9   |      |
| 2 단계   | 은행상승하락 | 0      | 22 | 14     | 61.1 |
|        |        | 1      | 13 | 22     | 62.9 |
| 전체 백분율 |        |        |    | 62.0   |      |
| 3 단계   | 은행상승하락 | 0      | 23 | 13     | 63.9 |
|        |        | 1      | 11 | 24     | 68.6 |
| 전체 백분율 |        |        |    | 66.2   |      |

a. 잘라낸 값은 .500임

〈표 6〉 은행 로지스틱 회귀분석 결과 방정식의 변수

|      |              | B      | S.E.   | Wald  | df | 유의수준 | Exp(B)                 |
|------|--------------|--------|--------|-------|----|------|------------------------|
| 1 단계 | 경기동행<br>순환변동 | -.669  | .348   | 3.695 | 1  | .055 | .512                   |
|      | 상수           | 66.806 | 34.775 | 3.691 | 1  | .055 | 1.0317134675799439E+31 |
| 2 단계 | 경기동행<br>순환변동 | -.794  | .358   | 4.908 | 1  | .027 | .452                   |
|      | 소비자기대        | .057   | .030   | 3.587 | 1  | .058 | 1.059                  |
| 3 단계 | 중국주가<br>상해   | -.002  | .001   | 3.749 | 1  | .053 | .998                   |
|      | 경기동행<br>순환변동 | -.912  | .378   | 5.841 | 1  | .016 | .402                   |
|      | 소비자기대        | .095   | .038   | 6.378 | 1  | .012 | 1.100                  |
|      | 상수           | 85.107 | 36.831 | 5.340 | 1  | .021 | 9.154E+36              |

### 3.2.3 증권

로지스틱 회귀 모형 :

$$p = 1 / (1 + \text{EXP}(-(0.261 * \text{한국설비물가지수} - 0.268 * \text{일본산업생산지수} + 0.001$$

$$* \text{니케이종합주가지수} - 0.003 * \text{상해종합주가지수} - 0.032 * \text{원화} + 12.211 * \text{위안화} - 0.180$$

$$* \text{엔화} + 0.000 * \text{금융기관유동성지수} + 5.176 * \text{장단기금리차} - 2.895 * \text{EuroTresury}$$

$$+ 207.424)))$$

38개의 변수들 중 설명력이 높은 변수를 채택하기 위해 후진단계 선택법으로 변수 채택한 결과, 총 28단계를 거쳐 채택된 변수는 한국설비물가, 일본산업생산지수, 중국산업생산지수, 일본주가, 니케이, 원화, 위안화, 엔화, 금융기관유동, 장단기금리차, EuroTress이다. 일본산업생산지수, 중국산업생산지수, 원화, 엔화, EuroTress는 증권 ETF 가격에 부의 영향을 준다. 반면, 한국설비물가, 일본주가, 니케이, 위안화, 금융기관유동, 장단기금리차는 주가 상승에 영향을 끼친다. 증권시장은 다양한 화폐의 흐름에 영향을 받는 것으로 나타났으며 가장 큰 영향은 원화의 변동에 따른 변동폭이 가장 큰 것으로 나타났다.

〈표 7〉 증권 로지스틱 회귀분석 결과 분류표

| 관측번호   |        | 예측값      |          | 정정 백분율 |      |
|--------|--------|----------|----------|--------|------|
|        |        | 증권상승하락 0 | 증권상승하락 1 |        |      |
| 27 단계  | 증권상승하락 | 0        | 26       | 11     | 70.3 |
|        |        | 1        | 6        | 27     | 81.8 |
| 전체 백분율 |        |          |          | 75.7   |      |
| 28 단계  | 증권상승하락 | 0        | 26       | 11     | 70.3 |
|        |        | 1        | 6        | 27     | 81.8 |
| 전체 백분율 |        |          |          | 75.7   |      |

a. 잘라낸 값은 .400임

〈표 8〉 증권 로지스틱 회귀분석 결과 방정식의 변수

|       |           | B       | S.E.   | Wald   | df | 유의수준 | Exp(B)     |
|-------|-----------|---------|--------|--------|----|------|------------|
| 27 단계 | 한국설비물가    | .293    | .103   | 8.060  | 1  | .005 | 1.340      |
|       | 일본산업생산    | -.294   | .106   | 7.687  | 1  | .006 | .745       |
|       | 일본주가니케이   | .001    | .001   | 3.311  | 1  | .069 | 1.001      |
|       | 중국주가상해    | -.003   | .002   | 2.937  | 1  | .087 | .997       |
|       | 원화        | -.037   | .011   | 11.831 | 1  | .001 | .964       |
|       | 위안화       | 11.480  | 5.133  | 5.002  | 1  | .025 | 96733.802  |
|       | 엔화        | -.213   | .106   | 3.994  | 1  | .046 | .808       |
|       | 금융기관유동    | .000    | .000   | 7.737  | 1  | .005 | 1.000      |
|       | 서비스업생산    | -.742   | .551   | 1.812  | 1  | .178 | .476       |
|       | 장단기금리차    | 4.887   | 2.207  | 4.905  | 1  | .027 | 132.566    |
| 28 단계 | EuroTrres | -3.023  | .950   | 10.125 | 1  | .001 | .049       |
|       | 상수        | 281.199 | 95.749 | 8.625  | 1  | .003 | 1.327E+122 |
|       | 한국설비물가    | .261    | .097   | 7.256  | 1  | .007 | 1.298      |
|       | 일본산업생산    | -.268   | .100   | 7.176  | 1  | .007 | .765       |
|       | 일본주가니케이   | .001    | .001   | 2.891  | 1  | .089 | 1.001      |
|       | 중국주가상해    | -.003   | .001   | 3.264  | 1  | .071 | .997       |
|       | 원화        | -.032   | .010   | 10.816 | 1  | .001 | .969       |
|       | 위안화       | 12.211  | 5.069  | 5.803  | 1  | .016 | 200914.223 |
|       | 엔화        | -.180   | .101   | 3.196  | 1  | .074 | .835       |
|       | 금융기관유동    | .000    | .000   | 8.119  | 1  | .004 | 1.000      |
|       | 서비스업생산    | 5.176   | 2.205  | 5.511  | 1  | .019 | 177.008    |
|       | EuroTrres | -2.895  | .909   | 10.151 | 1  | .001 | .055       |
|       | 상수        | 207.424 | 72.301 | 8.231  | 1  | .004 | 1.211E+90  |

## 3.2.4 철강

로지스틱 회귀 모형 :

$$p = 1/(1 + \text{EXP}(-(0.330 * \text{중국산업생산지수} + 0.375 * \text{구인구직비율} + 0.096 * \text{WTI} - 45.269)))$$

38개의 변수들 중 설명력이 높은 변수를 채택하기 위해 전진단계 선택법으로 변수 채택한 결과, 채택된 변수는 중국산업생산지수, 구인구직비율, WTI이다. 중국산업생산지수, 구인구직비율, WTI는 철강의 ETF 주가에 양의 영향을 끼친다. 즉 중국산업생산지수, 구인구직비율, WTI가 1증가 할 때마다 철강의 ETF주가는 각각 1.391배, 1.454배, 1.101배 상승할 가능성이 생긴다고 할 수 있다. 서부 텍사스산 원유가인 WTI가 철강업에 영향을 미치는 이유는 일반적으로 유가가 하락하면 제품가격이 하락하여 수익성이 떨어질뿐더러, 건설경기가 함께 하강하여 철강관련 건축자재의 수요가 줄어들게 되기 때문이다.

〈표 9〉 철강 로지스틱 회귀분석 결과 분류표

| 판측번호   |    | 예측값  |      |      | 정정 백분율 |  |
|--------|----|------|------|------|--------|--|
|        |    | 철강   |      | 1.00 |        |  |
|        |    | .00  | 1.00 |      |        |  |
| 1 단계   | 철강 | .00  | 17   | 7    | 70.8   |  |
|        |    | 1.00 | 14   | 16   | 53.3   |  |
| 전체 백분율 |    |      |      |      | 61.1   |  |
| 2 단계   | 철강 | .00  | 19   | 5    | 79.2   |  |
|        |    | 1.00 | 14   | 16   | 53.3   |  |
| 전체 백분율 |    |      |      |      | 64.8   |  |
| 3 단계   | 철강 | .00  | 20   | 4    | 83.3   |  |
|        |    | 1.00 | 11   | 19   | 63.3   |  |
| 전체 백분율 |    |      |      |      | 72.2   |  |

a. 잘라낸 값은 .600임

〈표 10〉 철강 로지스틱 회귀분석 결과 방정식의 변수

|      |         | B       | S.E.   | Wald  | Df | 유의수준 | Exp(B) |
|------|---------|---------|--------|-------|----|------|--------|
| 1 단계 | 중국-산업생산 | .215    | .108   | 3.966 | 1  | .046 | 1.240  |
|      | 상수      | -23.718 | 12.003 | 3.904 | 1  | .048 | .000   |
| 2 단계 | 중국-산업생산 | .280    | .128   | 4.753 | 1  | .029 | 1.323  |
|      | WTI     | .072    | .034   | 4.503 | 1  | .034 | 1.074  |
|      | 상수      | -37.537 | 15.771 | 5.665 | 1  | .017 | .000   |
| 3 단계 | 중국-산업생산 | .330    | .154   | 4.603 | 1  | .032 | 1.391  |
|      | 구인구직    | .375    | .171   | 4.808 | 1  | .028 | 1.454  |
|      | WTI     | .096    | .039   | 6.074 | 1  | .014 | 1.101  |
|      | 상수      | -45.269 | 19.023 | 5.663 | 1  | .017 | .000   |

다음의 표는 각 산업별 선택된 독립변수를 정리한 것이다.

〈표 11〉 각 산업별 선택된 독립변수

| 독립변수           | 반도체 | 은행 | 증권 | 철강 |
|----------------|-----|----|----|----|
| 한국-수출물가지수      |     |    |    |    |
| 한국-설비물가지수      |     |    | o  |    |
| 미국-필라델피아 연준 지수 |     |    |    |    |
| 미국-시카고 PMI 지수  |     |    |    |    |
| 미국-소비자 신뢰지수    |     |    |    |    |
| 일본-산업생산지수      |     |    | o  |    |
| 중국-소매판매지수      |     |    |    |    |
| 중국-산업생산지수      |     |    |    | o  |
| 미국주가지수-S&P500  |     |    |    |    |
| 일본주가지수-니케이225  |     |    | o  |    |
| 중국주가지수-상해종합    |     | o  | o  |    |
| 원화             |     |    | o  |    |
| 위안화            |     |    | o  |    |
| 유로             |     |    | o  |    |
| 엔화             |     |    |    |    |
| 경기선행지수         | o   |    |    |    |
| 경기동행지수         | o   |    |    |    |
| 경기동행지수 순환변동치   | o   | o  |    |    |
| 재고순환지표         |     |    |    |    |
| 소비자기대지수        |     | o  |    |    |
| 건설수주액          |     |    |    |    |

〈표 11〉 각 산업별 선택된 독립변수 (계속)

| 독립변수                                | 반도체 | 은행 | 증권 | 철강 |
|-------------------------------------|-----|----|----|----|
| 금융기관유동성                             |     |    | 0  |    |
| 산업생산지수                              |     |    |    |    |
| 제조업 가동율 지수                          | 0   |    |    |    |
| 건설 기성액                              |     |    |    |    |
| 서비스업 생산지수                           |     |    |    |    |
| 생산 자체품 재고지수                         |     |    |    |    |
| 기계수주액                               |     |    |    |    |
| 자본재수입액                              |     |    |    |    |
| 장단기금리차                              | 0   |    | 0  |    |
| 순상품교역조건                             |     |    |    |    |
| 구인구직비율                              |     |    |    | 0  |
| 경기선행지수 OECD(Amplitude Adjusted)     |     |    |    |    |
| 경기선행지수 OECD(Trend Restored)         |     |    |    |    |
| 경기선행지수 OECD(Europe, Trend Restored) |     |    |    |    |
| 금                                   |     |    |    |    |
| WTI                                 |     |    |    | 0  |
| 브랜트유                                |     |    |    |    |

다음은 각 산업별 도출된 회귀모형에 대한 식이다.

〈표 12〉 각 산업별 회귀모형

| 업종  | 회귀모형  |
|-----|---|
| 반도체 | $p = 1/(1 + EXP(-(-1.035 * 경기선행 + 1.466 * 경기동행 - 2.822 * 경기동행순환변동 + 0.266 * 제조업가동율 + 5.057 * 장단기금리차 + 207.719)))$   |
| 은행  | $p = 1/(1 + EXP(-(-0.002 * 상해종합주가지수 - 0.912 * 경기동행순환변동 + 0.095 * 소비자기대지수 + 85.107)))$   |
| 증권  | $p = 1/(1 + EXP(-(0.261 * 한국설비물가지수 - 0.268 * 일본산업생산지수 + 0.001 * 니케이종합주가지수 - 0.003 * 상해종합주가지수 - 0.032 * 원화 + 12.211 * 위안화 - 0.180 * 엔화 + 0.000 * 금융기관유동성지수 + 5.176 * 장단기금리차 - 2.895 * EuroTreasury + 207.424)))$ |
| 철강  | $p = 1/(1 + EXP(-(0.330 * 중국산업생산지수 + 0.375 * 구인구직비율 + 0.096 * WTI - 45.269)))$  |

### 3.3 예측

3.2에서 구축한 로지스틱 회귀 모형을 이용하여 2015년 12개의 월별자료로 구성된 예측용 자

료에 대해 ETF의 상승 혹은 하락을 예측해보았다. 예측용 자료의 입력 변수들의 값이 주어졌을 때 모형화 자료를 통해 추정된 로지스틱 회귀 모형으로 종속변수가 1이 될 확률, 즉 ETF가 상승 할 사후 확률을 구할 수 있다. 사후 확률이 절단값 보다 높으면 종속변수를 1(상승)로 분류하고, 절단값 보다 낮으면 0(하락)으로 분류하였다. 이렇게 분류한 종속변수와 실제 상승, 하락 여부와 비교하여 모형이 얼마나 잘 예측하였는지 예측 정확도를 계산할 수 있다.

〈표 13〉 모형 정확도 및 예측 정확도

|     | 모형 정확도 | 예측 정확도 |
|-----|--------|--------|
| 반도체 | 0.775  | 0.667  |
| 은행  | 0.662  | 0.583  |
| 증권  | 0.722  | 0.416  |
| 철강  | 0.757  | 0.75   |

위의 표에서 예측용 자료의 정확도는 증권을 제외한 반도체, 은행, 철강의 경우에는 거의 비슷한 결과를 얻음을 확인할 수 있다. 증권의 경우에는 정확도가 40%가 넘게 하락하였는데, 2013~2014년에 통과된 국내 투자은행의 활성화, 자산운용산업 규제체계의 선진화의 내용을 포함한 자본시장법 개정과, 정부가 발표한 금융업 경쟁력 강화 방안 등의 변화를 감안하면 이는 2015년 이전과 달라진 증권시장의 환경에 기인한다고 생각된다.

#### IV. 결 론

본 연구는 각 산업별 ETF 주가지수를 국내, 해외 경제지표를 바탕으로 예측하였다. 로지스틱 회귀 분석 방법을 이용하여 유의미한 변수들을 추출하여서 모형을 만들고 그를 통하여 주가를 예측해보는 시도도 하였다. 쇠익선 등이 진행한 선행연구의 예측정확도가 약 60%인 것과 비교했을 때 본 연구의 평균적인 예측 정확도도 60%인 것으로 보아 비슷한 방식을 통한 로지스틱 회귀모형은 시기와 큰 관계없이 비슷한 정확도를 갖고 있음을 살펴볼 수 있었다.

로지스틱 회귀 분석을 토대로 각 산업별 ETF 주가지수에 영향을 미치는 변수를 선택해본 결과, 반도체ETF주가는 경기동행, 경기동행순환변동, 제조업가동율, 장단기금리차에 영향을 받으며 은행ETF주가는 은 상해종합주가지수, 경기동행순환변동, 소비자기대지수에 영향을 받는 것으로 나타났다. 증권업은 한국설비물가지수, 일본산업생산지수, 니케이종합주가지수, 상해종합주가지수, 원화 위안화, 엔화, 금융기관유동성지수, 장단기금리차에 유의미한 영향을 받는 것으로 나타났으

며 철강 ETF주가는 중국산업생산지수, 구인구직비율이 큰 영향을 미쳤다. 반도체를 제외한 3개의 산업 군에서는 중국이 국내 산업에 큰 영향을 끼친다는 것을 확인할 수 있다. 또한 반도체 산업의 경우에는 해외변수에 별로 의존하지 않고 국내 경제지표에 따라 움직이는 것을 확인할 수 있었다. 이에 반해 증권업은 다양한 해외변수에 영향을 받는다는 것을 확인할 수 있었다.

본 연구는 빅데이터 분석방법을 이용하여 종합주가지수가 아닌 산업별 ETF 주가지수를 예측해 본다는 점에서 의의가 있다. 이러한 비즈니스 분석을 바탕으로 각 산업별 ETF 주가를 파악하여 국내외 경제 상황에 따라 각 업종이 상승, 또는 하락할지에 대한 정보를 얻을 수 있다. 이러한 연구결과를 통해 일반 투자자에게는 어느 업종에 투자해야 하는지에 대한 투자정보를 줄 수 있으며 해당 업종의 경영자들에게는 통계적인 예측을 기반으로 한 경제 환경의 변화에 따른 기업 경영의 유연화를 도모하는 방향으로 변화를 모색할 수 있다. 본 연구의 한계점은 ETF를 가지고 있는 업종이 제한적이어서 다양한 산업군에 대한 분석을 해보기 어려웠으며, ETF가 비교적 최근에 만들어진 상품으로 충분치 못한 과거 데이터만을 제한적으로 이용했다는 것이다. 또한 모델링을 위한 분석용 자료 이외에 검증용 자료를 통한 모델의 검증이 필요했다.

### 참 고 문 헌

1. 고희운, 강상훈 (2016). 글로벌 금융위기와 국제원유의 충격이 산업별 주가지수 변동성에 미치는 영향. 2016 재무금융 관련 5개 학회 학술연구발표회, p.2316-2337.
2. 김삼용, 김진아 (2009). 일반 자기회귀 이분산 모형을 이용한 시계열 자료 분석. Journal of Korean Data & Information Science Society 2009, 20(3), p.475-483.
3. 육근철, 김찬규, 김영희 (1999). 인공 신경망을 이용한 KOSPI 200 지수의 상향식 예측모형에 관한 연구. 대한산업공학회/한국공업경영학회 '99 추계공동학술대회 논문집, Session A17.3 인공지능 I.
4. 이형용 (2008). 한국 주가지수 등락 예측을 위한 유전자 알고리즘 기반 인공지능 예측기법 결합모형. Entrue Journal of Information Technology, Vol. 7, p.33-43.
5. 이희춘 (2004). KODEX200 주가 수익 예측에 대한 연구. 경영교육논총, 제 37집, p.345-358.
6. 최익선, 강동식, 이정호, 강민우, 송다영, 신서희, 손영숙 (2012). 국내외 경제지표를 예측변수로 사용한 산업별 주가지수 예측. Journal of Korean Data & Information Science Society 2012, 23(2), p.271-283

7. Balkishan Sharma, Ravikant Jain (2014). Right choice of a method for determination of cut-off values: A statistical tool for a diagnostic test. *ASIAN JOURNAL OF MEDICAL SCIENCES 2014*.
8. Stephan Dreiseitl, Lucila Ohno-Machado (2002). Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Information*, Volume 35, Issues 5–6, p.352–359