

## 국내외 경제지표를 예측변수로 사용한 산업별 주가지수 예측

Prediction of the industrial stock price index using domestic and foreign economic indices

---

저자 (Authors)	최익선, 강동식, 이정호, 강민우, 송다영, 신서희, 손영숙 Ik Sun Choi, Dong Sik Kang, Jung Ho Lee, Min Woo Kang, Da Young Song, Seo Hee Shin, Young Sook Son
출처 (Source)	<a href="#">한국데이터정보과학회지 23(2)</a> , 2012.4, 271-283(13 pages) <a href="#">Journal of the Korean Data And Information Science Society 23(2)</a> , 2012.4, 271-283(13 pages)
발행처 (Publisher)	<a href="#">한국데이터정보과학회</a> The Korean Data and Information Science Society
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07244708">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07244708</a>
APA Style	최익선, 강동식, 이정호, 강민우, 송다영, 신서희, 손영숙 (2012). 국내외 경제지표를 예측변수로 사용한 산업별 주가지수 예측. 한국데이터정보과학회지, 23(2), 271-283
이용정보 (Accessed)	아주대학교 202.30.7.*** 2020/02/18 14:21 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# 국내외 경제지표를 예측변수로 사용한 산업별 주가지수 예측<sup>†</sup>

최익선<sup>1</sup> · 강동식<sup>2</sup> · 이정호<sup>3</sup> · 강민우<sup>4</sup> · 송다영<sup>5</sup> · 신서희<sup>6</sup> · 손영숙<sup>7</sup>

<sup>1234567</sup> 전남대학교 통계학과

접수 2012년 1월 30일, 수정 2012년 3월 1일, 게재확정 2012년 3월 6일

## 요 약

본 연구에서는 모든 산업을 총합한 종합주가지수 예측을 다루는 기존의 연구들과는 달리 11개의 대표 산업별 주가지수의 상승 및 하락을 예측하였다. 해외경제상황에 큰 영향을 받는 우리나라 주식 시장을 고려하여 국내 경제지표뿐만 아니라 미국, 일본, 중국, 유럽의 주요 경제지표를 예측변수로 사용하였다. 2001년부터 2011년까지 총 132개의 월별 자료에 대하여 로지스틱 회귀모형과 신경망모형에 의한 분석은 대체로 60% 내외의 정확도를 보였다.

주요용어: 국내외 경제지표, 로지스틱 회귀모형, 산업별 주가지수 예측, 신경망모형.

## 1. 서론

최근 들어 한국의 주식시장은 예측하기 힘들만큼 변동성이 심화되고 있는데 그 이유 중 하나로서 다른 아시아 국가에 비해 외국인 비중이 높기 때문이다. 2011년 9월 한 달간 유로존 위기 심화로 외국인 자금의 이탈 강도가 거세졌고 외국인인 코스닥을 포함한 국내 증시에서 순매도액만 2조원을 넘어섰다. 미국발 위기가 몰아치던 8월부터 9월까지 순매도액은 7조1650억에 이르렀고, 증시는 200포인트 이상의 하락폭을 보였다. 2011년 9월 현재, 국내 증시에서 외국인 비중은 32%로 인도 (19%), 인도네시아 (17%), 중국 (9%) 등에 비해 상당히 높은 수치다. 그러므로 예측하기 힘든 외국인의 자본의 유입, 유출로 인하여 국내 경제 상황만으로는 기업을 판단하고 주식에 투자하는 것은 어려울 것으로 판단된다. 한국의 기업 상황이 아무리 호황기라도 해외 경제 상황의 영향을 받지 않을 수 없다. 한국은 2010년의 경제회복국면을 거쳐 2011년에도 계속적인 수출호재와 기업의 이익증대로 인해 전반적인 기업들의 주가는 상승하였고, 종합주가지수 (Korea composite stock price index; KOSPI)는 최고가를 갱신하면서 시가총액은 1,000조를 넘어섰으며 지수는 2,200을 넘어섰다. 하지만 그리스의 지속적인 재정 적자 상황이 2010년부터 유럽 대부분 국가들의 신용부도스왑 (Credit default swap; CDS) 프리미엄을 급격히 상승시켰다. 결국, 유럽에서 경제규모 3위인 이탈리아도 신용등급이 강등되었다. 유럽의 채무불이행은 전 세계의 주식시장과 금융시장을 불안하게 만들었고, 2011년 8월에 결국 미국과 일본의 신용등급 강등과 경제지표의 악화는 상승장구하던 KOSPI 지수를 쉽게 무너뜨리고 말았다. 이제 투자자라면 해외경제의

<sup>†</sup> 이 논문은 2010년도 전남대학교 학술연구비 지원에 의하여 연구되었음.

<sup>1</sup> (500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 학부졸업생.

<sup>2</sup> (500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 학부생.

<sup>3</sup> (500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 학부생.

<sup>4</sup> (500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 학부생.

<sup>5</sup> (500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 학부졸업생.

<sup>6</sup> (500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 학부생.

<sup>7</sup> 교신저자: (500-757) 광주광역시 북구 용봉동 300, 전남대학교 통계학과, 교수. E-mail: ysson@jnu.ac.kr

상황을 고려하지 않을 수 없다. 또한 산업마다 영향을 받는 정도와 관련된 변수가 각각 다르다. 예를 들면, 수출산업은 국제 경제상황에 민감하고, 내수산업은 국내 경제상황에 민감할 것이다.

KOSPI는 국내외 정치 및 경제를 포함한 예측 불가능한 여러 요인들을 포함하고 있기 때문에 KOSPI의 정확한 단기 예측은 상당히 어려운 문제에 속한다. 그럼에도 불구하고 KOSPI는 랜덤하지 않고 설명하기는 어렵지만 어떤 패턴이 있다고 가정하고 여러 모형들에 의해 예측이 이루어져 왔다. 모든 산업을 총합한 KOSPI의 예측 및 분석을 다루는 기존의 연구들은 다음과 같다.

김유일 등 (2004)은 KOSPI 200지수와 S&P 500지수의 상승 및 하락을 예측하기 위하여 신경망모형과 SVM (Support vector machine) 기법을 사용하였다. 예측변수로서는 가격변화를 나타내는 지표, 주가의 전환점을 알려주는 오실레이터 지표, 장기간의 추세 경향을 나타내는 지표, 기복이 심한 변수의 움직임 완만하게 표현해내는 이동평균 지표 등 11개의 기술적 지표들을 사용하였다. 구간을 달리 하는 KOSPI 200 지수의 5개의 자료세트에 대하여 신경망모형은 55%-59%, SVM은 55%-58%의 정확률을 보였다.

Lee (2008)는 Kim (2003)에서 사용된 12개의 기술적 지표를 예측변수로 사용하고, 유전자 알고리즘 기반하의 인공지능 예측기법 결합모형을 이용하여 로지스틱 회귀모형, 신경망모형, 그리고 SVM모형의 예측 결과들을 결합하여 KOSPI 지수 등락을 예측하였다. 결합모형에 의한 예측은 최고 66%까지의 정확도를 보였다.

박주환과 정진혁 (2008)은 KOSPI의 시가, 고가, 저가, 종가 4가지의 데이터로 구성된 16개의 기술적 지표를 예측변수로 사용하고, SVM모형으로 예측한 일간예측은 56.6%, 주간예측은 56.3%의 적중률을 나타냈다. 모형의 예측에 맞추어서 매수 혹은 매도 포지션을 취한다고 가정했을 때 수익률 시뮬레이션 결과는 일간 모형 (4,600영업일, 약 17년)의 경우 세금을 감안하더라도 3,000%에 달하는 수익률을 기록하였다.

이은진 등 (2008)은 신경망모형을 사용하여 KOSPI 지수를 예측하였다. KOSPI, KOSPI 200, DOW, NASDAQ의 거래량, 거래대금, 시가, 고가, 저가, 종가, 그리고 원달러 환율, 등, 총 22개의 입력변수 중에서 먼저 주성분분석에 의하여 12개의 변수를 추출하였다. 추출된 12개 입력변수에 대하여 예측하고자 하는 예측일에 따라 최량부분적합법을 이용하여 최종적으로 예측변수들을 선정하였다. 성능평가를 위해 주가지수의 변동폭이 다른 두 종류의 실험데이터를 대상으로 예측을 진행한 결과 30일의 연속적인 KOSPI 지수예측에 있어 11.92포인트의 평균오차를 보였다.

김삼용과 김진아 (2009)는 KOSPI 수익률에 대한 최적의 모형을 구축하기 위하여 다양한 형태의 조건부 평균모형과 조건부 분산모형을 적합시켰으며, 박인찬 등 (2009)은 KOSPI 200 지수의 전일대비 상승 혹은 하락을 예측하기 위하여 시계열 요소분해모형과 ARIMA 모형을 사용하였다.

신동근과 정경용 (2011)은 가중 퍼지소속함수를 이용한 신경망모형에 의해서 5일간의 KOSPI 단기 추세 예측을 위하여 비중복면적 분산 측정법을 사용하여 여러 개의 특징입력 중에서 중요도가 가장 낮은 특징입력을 자동적으로 하나씩 제거하면서, 최소화된 특징입력으로 최대의 예측성능을 찾았다. 실험결과에서 5일 간의 주가 단기추세의 상승과 하락에 대한 73.84%의 정확도를 보여주었다.

한나영 (2011)은 2006년부터 2010년까지 KOSPI 일별자료에 대해 기존의 연구와는 다르게 2008년 국제금융위기 시기를 기준으로 금융위기 발생 이전과 발생 이후의 주가를 예측하기 위하여 Kim (2003)에서 사용된 12개의 기술적 지표를 예측변수로 선정하였고, 신경망모형, 이동평균법, 지수평활법, 그리고 다중회귀모형을 사용하여 예측하였다. 최종 예측성능을 RMSE (Root mean square error)로 비교한 결과 신경망모형이 다른 모형들보다도 더 높은 성과를 보인 것으로 파악되었다.

불안정한 금융시장의 변동성을 분석하기 위하여 오경주 등 (2011)은 선형 및 신경망 자기회귀모형을 이용하여 주식시장의 불안정지수를 개발하였고, 박범조 (2011)는 개별 주가에 반영된 시변 무리행동을 연구하였다.

모든 산업을 총합한 KOSPI 예측을 다루는 기존의 연구들과는 달리 본 연구의 목적은 11개의 대표 산업별 주가지수의 전월 대비 상승 혹은 하락을 예측하는 것이다. 각 산업의 주가지수를 예측하기 위하여 국내변수 (국내경제지표)와 해외변수 (미국, 일본, 중국, 유럽의 주가와 주요 경제지표) 들 중에서 각 산업마다 민감하게 작용하는 입력변수를 예측변수로 사용함으로써 산업별로 예측의 정확도를 높이고자 하였다. 이는 국내외 경제상황에 따라 주식투자자들이 어느 산업에 투자해야 하는지의 투자정보를 줄 수 있다는 점에 의의를 둔다.

수치분석을 위하여 2001년부터 2011년까지의 11년 동안에 걸친 각 산업주가지수의 132개 월별자료를 목표변수 자료로 사용하였고, 총 39개의 국내외 경제변수들 중에서 단계별 변수선택법 (Stepwise variable selection)에 의해 추출된 예측변수들 위에서 로지스틱 회귀모형과 신경망모형에 의하여 각 산업별 주가지수의 전월대비 상승 혹은 하락을 예측하였다. 모든 자료분석은 SAS E-Miner를 사용하여 수행되었다.

## 2. 자료설명

목표변수로 사용된 표 2.1의 11개 대표 산업별 주가지수는 KOSPI를 대표할 수 있는 산업으로서 거래량을 기준으로 거래량이 많은 상위 11개 산업을 선택하였다. 표 2.2에 보이는 입력변수들 중에서 국내변수로는 설비투자지수, 경기선행지수, 경기동행지수, 산업생산지수 등을, 그리고 해외변수로는 미국, 유럽, 중국, 일본의 주가지수, 환율, 원유가격, 및 금가격 등을 1개월 전 변동률로 변환하여 총 38개의 변수를 기본이 되는 입력변수로 구성하였다. 또한 각 변수의 3, 6, 12개월 전 변동율도 입력변수 세트에 포함시켰으나 변수선택과정에서 대부분 선택되지 않거나 선택되더라도 낮은 정확도를 보이거나, 민감도 혹은 특이도가 0이 되어 3, 6, 12개월 전 변동율들은 입력변수 세트에서 제외되고 유일하게 선택된 경기선행지수의 12개월 전 변동율만 추가되어 총 39개의 입력변수의 세트를 표 2.2와 같이 구성하였다.

표 2.1 목표변수 설명과 각 자료별 상승 (1)의 구성 비율

목표변수	산업설명	분석용자료	검증용자료
y1	제조산업	0.467	0.689
y2	기계산업	0.533	0.689
y3	건설산업	0.511	0.644
y4	철강산업	0.622	0.600
y5	음식료산업	0.489	0.667
y6	전기전자산업	0.444	0.600
y7	운수창고산업	0.444	0.556
y8	금융산업	0.467	0.644
y9	유통산업	0.511	0.600
y10	서비스산업	0.467	0.689
y11	화학산업	0.578	0.711

(\*) 각 목표변수는 산업별 주가지수의 전월대비 상승의 경우는 1의 값을, 하락의 경우는 0의 값을 가지는 이항자료이다.

수치분석을 위하여 사용된 자료는 2001년부터 2011년까지의 132개 월별자료이다. 월별자료를 사용한 이유는 입력변수로 사용될 경기선행지수, 소비자자기대지수, 그리고 산업생산지수 등 대부분의 국내의 경제지표가 한 달에 한번 발표되는 것이 많으므로 주가지수도 월별자료를 사용하였다. 이때, 주가지수의 월간자료는 매달 주식거래 첫날의 증가를 의미한다. 11개의 대표 산업별 주가지수의 전월 대비 상승 (1), 혹은 하락 (0)의 값을 갖는 이범주 자료를 목표변수로 설정하였다.

총 132개의 자료들 중에서 2001년 1월부터 2008년 6월까지의 90개 자료는 모형화를 위한 자료로 사용하였고, 2008년 7월부터 2011년 12월까지의 42개 자료는 예측용 자료 (Test data)로 사용하였다. 이

때 모형화를 위한 90개의 자료 중 각 45개 씩을 분석용 자료 (Training data)와 검증용 자료 (Validation data)로 랜덤추출하여 사용하였다. 각 산업별 주가지수의 전월 대비 상승 (1)의 구성 비율이 분석용 및 검증용 자료별로 표 2.1에 나타나 있다.

표 2.2 국내의 입력변수 (전월대비 변동율)

입력변수	변수설명	입력변수	변수설명
x1	한국-수출 물가지수	x21	소비자기대지수
x2	한국-설비투자지수	x22	건설주주액
x3	미국-필라델피아 연준 지수	x23	금융기관유동성
x4	미국-시카고 PMI지수	x24	산업생산지수
x5	미국-소비자신뢰지수	x25	제조업 가동율 지수
x6	일본-산업생산지수	x26	건설 기성액
x7	중국-소매판매지수	x27	서비스업 생산지수
x8	중국-산업생산지수	x28	생산자제품 재고지수
x9	미국주가지수-다우산업	x29	기계주주액
x10	일본주가지수-니케이 225	x30	자본재수입액
x11	중국주가지수-상해종합	x31	장단기금리차
x12	원화	x32	순상품교역조건
x13	위안화	x33	구인구직비율
x14	유로	x34	경기선행지수 OECD (Amplitude Adjusted)
x15	연화	x35	경기선행지수 OECD (Trend Restored)
x16	경기선행지수 12	x36	경기선행지수 OECD (Europe, Trend Restored)
x17	경기선행지수	x37	금
x18	경기동행지수	x38	WTI
x19	경기동행지수 순환 변동치	x39	브렌트유
x20	재고순환지표		

(\*) x16 (경기선행지수 12)는 경기선행지수의 12개월 전대비 변동율을 의미한다.

목표변수와 입력변수의 모든 원자료의 출처는 인포맥스 (<http://www.infomax.co.kr>)이고, 원자료는 모두 변동률로 변환하여 사용하였다.

### 3. 자료분석

목표변수인 산업별 주가지수의 전월대비 상승 (1) 혹은 하락 (0)을 예측하기 위하여 11개 산업 각각에 대해 SAS E-Miner를 사용하여 로지스틱 회귀분석과 신경망분석을 실시하였다. 본 논문에서는 11개 산업 중에서 화학산업 (y11)에 대한 자료분석 과정 및 결과에 대한 설명을 상세히 하고 나머지 산업에 대한 분석결과는 요약하여 설명하고자 한다.

#### 3.1. 변수선택

로지스틱 회귀분석 혹은 신경망분석을 위하여 표 2.2에 제시되어 있는 39개의 국내외 경제변수를 모두 예측변수로 사용하기에는 자료 수에 대비하여 매우 많다. 따라서 SAS E-Miner에서 R-square를 변수선택기준으로 한 Variable selection 노드를 이용하여 영향력 있다고 생각되는 유의한 10 여개의 변수를 선택한 후, 선택된 변수들에 대하여 로지스틱 회귀모형의 적합결과를 통해 유의수준 15% 내외에서 유의하다고 생각되는 예측변수들을 최종 선택하기로 한다.

R-square를 변수선택 기준으로 한 Variable selection 노드는 다음과 같이 입력변수의 선택을 한다. 즉, 각 입력변수와 목표변수와의 단순 상관계수의 제곱이 기준값 (디폴트 기준값은 0.005)보다 작은 입력변수를 제거한 후, 제거되지 않고 남아있는 변수들 중에서 전진 단계별 R-square 회귀변수선택법을

적용하여 R-square 증가분이 기준값 (디폴트 기준값은 0.0005)보다 작은 변수가 나타나면 변수선택을 중단한다.

표 3.1은 화학산업의 예측변수 선택결과를 나타낸 표로서 각 변수 선택단계에서 선택된 변수의 경우는 유의확률이 표시되어 있다. 입력변수의 수를 줄이기 위하여 1차적으로 R-square를 변수선택 기준으로 한 Variable selection 노드를 수행하여 유의한 변수 순으로 10개의 입력변수 (표 3.1의 3열)가 선택되었다. 이렇게 선택된 10개의 변수를 입력변수로 하여 로지스틱 회귀분석을 실시한 결과 유의수준 15%에서 6개의 변수 (표 3.1의 4열)가 유의하였다. 이제 6개의 변수를 화학산업의 주가지수 예측을 위한 최종 예측변수로 선택하기로 한다.

표 3.1 로지스틱회귀에 의한 변수선택 (목표변수: 화학산업)

입력변수	변수 설명	1차 변수선택	2차 변수선택
x2	한국-설비투자지수	0.018	0.081
x3	미국-필라델피아 연준 지수	0.039	0.074
x7	중국-소매판매지수	0.041	0.047
x9	주가지수-다우산업	0.000	0.001
x12	원화	0.048	
x15	엔화	0.058	
x26	건설 기성액	0.038	
x32	순상품교역조건	0.017	0.079
x35	경기선행지수 OECD (Trend Restored)	0.128	
x38	WTI 국제유가	0.000	0.004

화학산업에 대해 선택된 예측변수들을 유의한 순서대로 나열하면 다우존스 주가지수 (x9), WTI 국제유가 (x38), 중국-소매판매지수 (x7), 미국-필라델피아 연준지수 (x3), 순상품 교역조건 (x32), 그리고 한국-설비투자지수 (x2) 이다. 이렇게 선택된 예측변수들은 목표변수인 화학산업과 비교적 연관성 있는 변수들이라고 볼 수 있다. 서브프라임 모기지 사태로 시작된 미국발 금융위기가 세계 금융 시장을 뒤흔든 사례를 보면 미국이 국제 금융 시장에 미치는 영향은 매우 크다고 볼 수 있다. 미국 경제를 반영하는 다우존스 주가지수는 당연히 화학업은 물론 다른 산업별 주가지수에도 지대한 영향을 미칠 것으로 보인다. 화학산업의 주된 산업이 원유를 수입하여 정제 과정을 거쳐 만들어진 석유화학제품을 판매, 수출하는 산업이다. 화학산업의 주재료가 원유인 만큼 화학산업에서 두 번째로 유의한 예측변수인 WTI 국제유가는 화학산업과 매우 관련이 깊다고 할 수 있다. 화학산업의 경우 매출의 반 이상이 수출을 통해 이루어지는데 주로 중국, 일본에 정제된 석유화학제품을 수출하고, 미국에 휘발유를 수출한다. 따라서 중국 경기와 미국 경기 전망을 알 수 있는 중국 소매판매지수와 미국 필라델피아 연준 지수도 일정 부분 화학산업의 주가 지수에 영향을 미칠 것으로 보인다. 화학산업은 그 산업구조의 특성상 수출입 조건에 민감하게 반응하므로 순상품교역조건도 화학산업 지수에 연관성이 있다고 할 수 있다. 그러나 마지막으로 선택된 한국설비투자지수의 경우 화학산업과의 직접적인 연관성은 찾기 힘들다.

화학산업과 같은 방법으로 나머지 10개의 산업에 대해서도 각 산업별로 Variable selection 노드와 로지스틱 회귀분석을 통해 유의수준 15% 내에서 유의한 변수 순으로 최종적인 예측변수를 선택을 하였다. 표 3.2는 각 산업별로 5개 내외의 최종적인 예측변수를 선택한 결과와 그 때의 유의확률을 나타낸 표이다.

경제적인 관점에서 각 산업과 연관성 있는 예측변수들이 잘 선택되었는지를 살펴보기로 하자. 표 3.2에서 선택된 예측변수들을 살펴보면 다우존스 주가지수 (x9)의 경우 모든 산업에서 선택되었음을 볼 수 있다. 앞에서 언급했듯이 미국이 국제 금융 시장에 미치는 영향은 매우 크다고 볼 수 있다. 특히, 우리나라는 경제 자체의 규모는 작지만 자본시장에 있어 개방도가 높은 소규모 개방경제이다. 따라서 미국 경제를 반영하는 다우존스 주가지수는 당연히 우리나라 주가지수에도 지대한 영향을 미칠 것으로 보



인다. WTI 국제유가 (x38) 또한 기계산업을 제외한 모든 산업에서 선택된 것을 볼 수 있다. 이는 우리나라의 경우 자원이 부족한 나라로 원유의 경우 100% 수입하고 있는 실정이기에 국제유가의 경우 수입 원자재 가격에 영향을 미쳐 각 산업별 주가지수에도 큰 영향을 미칠 것으로 보인다.

다우존스 주가지수와 WTI 국제유가 이외에 각 산업별로 선택된 예측변수들을 살펴보면, 제조산업 주가지수 (y1)은 금융기관유동성 (x23)이 예측변수로 선택되었으나 연관성을 찾기 힘들다.

기계산업 주가지수 (y2)는 중국-소매판매지수 (x7), 니케이 225 주가지수 (x10), 자본재수입액 (x30)이 예측변수로 선택되었다. 국내의 중공업회사가 대량의 굴삭기를 중국에 수출하거나 조선회사가 플랜트 사업으로 해외에 수출하는 경향으로 보아, 중국의 성장을 나타내는 소매판매지수는 기계업과 관련이 있다. 또한 일본도 우리나라와 더불어 기계산업이 발달되어 있기 때문에 니케이 주가지수도 기계산업과 밀접한 관련이 있으며, 자원이 부족한 우리나라 특성상 원자재를 수입하는 자본재수입액도 그 연관성이 있다.

표 3.2 로지스틱회귀에 의한 산업별 최종 변수선택의 결과

	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10
	제조산업	기계산업	건설산업	철강산업	음식료산업	전기전자산업	운수창고산업	금융산업	유통산업	서비스산업
X3					0.072					
X4								0.067		
X5				0.001						
X6							0.091			
X7		0.135								
X9	0.001	0.002	0.003	0.097	0.001	0.001	0.001	0.004	0.001	0.065
X10		0.048						0.108		
X12					0.035					0.003
X14					0.085	0.092				0.085
X15										0.077
X16								0.028		
X23	0.130			0.098						
X27				0.061						
X28							0.127			
X30		0.033				0.051				
X32				0.009						
X33						0.001				
X34										0.004
X36										0.017
X38	0.001		0.057	0.012	0.008	0.002	0.116	0.028	0.059	0.006
x39								0.067		

철강산업 주가지수 (y4)는 미국-소비자신뢰지수 (x5), 금융기관유동성 (x23), 서비스업생산지수 (x27), 순상품교역조건 (x32)이 예측변수로 선택되었다. 철강산업에서 대표적인 자동차산업이 해외 시장으로의 수출에 크게 의존하는 산업이므로 순상품 교역조건, 미국-소비자 신뢰지수 등 수출입에 관련된 변수들과 강한 연관성이 있는 것을 보였다.

음식료산업 주가지수 (y5)는 미국-필라델피아 연준 지수 (x3), 원화 (x12), 유로 (x14)가 예측변수로 선택되었다. 음식료의 경우 원화, 유로 즉 환율과 밀접한 관계가 있는 걸로 나타났다. 이는 우리나라가 밀가루와 같은 대부분의 곡물원재료를 미국으로부터 수입하기 때문에 미국-필라델피아 연준지수와 환율의 영향을 많이 받는다고 할 수 있다.

전기전자산업 주가지수 (y6)는 유로 (x14), 자본재수입액 (x30), 구인구직비율 (x33)이 예측변수로 선택되었는데, 우리나라의 경우 대부분 원재료를 해외에서 수입으로 의존하므로 자본재수입액은 연관성이 있으나, 구인구직비율과 유로는 전기전자산업과 연관성을 찾기는 힘들다.

운수창고산업 주가지수 (y7)는 일본-산업생산지수 (x6), 생산자제품재고지수 (x28)이 예측변수로 선택되었다. 운수창고산업은 기본적으로 제품을 보관하고 운송하는 산업이므로 생산자제품재고지수와는

연관성이 있다. 또 우리나라의 경우 일본과 근접해있어 빈번하고 중요한 교역을 많이 하므로 일본 산업 생산지수와도 연관성이 있다.

금융산업 주가지수 ( $y_8$ )는 미국-시카고 PMI지수 ( $x_4$ ), 주가지수-니케이 225 ( $x_{10}$ ), 경기선행지수 12 ( $x_{16}$ ), 브렌트유 ( $x_{39}$ )가 예측변수로 선택되었다. 우리나라는 미국, 일본과의 교역이 많기 때문에 미국-시카고 PMI지수와 니케이 225 주가지수에 영향을 받는다. 특히 PMI지수는 경기에 선행하는 자료이므로 금융산업에는 영향을 미치게 된다. 또한 경기선행지수를 보고 투자와 거래가 이루어지므로 금융산업과 연관성을 갖는다. 그러나 브렌트유와 금융산업과의 관계는 연관성을 찾기 힘들다.

마지막으로 서비스산업 주가지수 ( $y_{10}$ )는 원화 ( $x_{12}$ ), 유로 ( $x_{14}$ ), 엔화 ( $x_{15}$ ), OECD 토탈 경기선행지수 ( $x_{34}$ ), OECD 유로 경기선행지수 ( $x_{36}$ )가 예측변수로 선택되었다. 서비스산업은 유통서비스업, 금융서비스업, 오락문화 서비스업, 통신서비스업 등을 모두 포괄하는 산업으로 경제 상황에 민감하게 반응한다. 따라서 OECD 토탈 경기선행지수와 OECD 유로 경기선행지수는 향후 경기를 예측하는 지표로 경제 상황에 민감하게 반응하는 서비스산업과 연관성이 있다고 할 수 있다. 또한 환율 (원달러, 엔화, 유로)은 우리나라 수출품 가격에 영향을 미쳐 소비를 감소시킬 수 있으므로 서비스산업에도 그 영향이 미칠 것으로 보인다.

목표변수와 선택된 예측변수들 간의 연관성을 살펴본 결과 각 산업별로 영향력 있는 변수들이 비교적 잘 선택된 것을 확인 할 수 있다. 이렇게 총 11개 산업에 대해 최종적으로 선택된 예측변수들을 사용하여 로지스틱 회귀분석과 신경망분석을 실시해 보았다.

### 3.2. 로지스틱 회귀분석

표 3.3은 화학산업에 대하여 최종 선택된 6개의 예측변수 위에서 로지스틱 회귀분석을 수행한 결과로서 각 절단값에 따른 분류의 정확도, 민감도, 특이도를 분석용 자료와 검증용 자료별로 보여준다. 정확도는 주가상승 (1) 및 주가하락 (0)을, 민감도는 주가상승 (1)을, 특이도는 주가하락 (0)을 정확하게 분류하는 척도이므로 정확한 예측을 위해서는 정확도, 민감도, 특이도 값들이 모두 높은 절단값을 선택하는 것이 좋다. 또한 모형을 추정하는 분석용 자료와 추정된 모형을 검증하는 검증용 자료에서의 통계값 결과들이 비슷한지를 검토하는 안정성 (Stability) 평가를 해야 한다. 이는 안정성이 보장된 예측모형에 의해서 향후 분석용 및 검증용 자료와 유사한 자료 구성을 가지는 예측용 자료에 대해서도 안정적인 예측을 할 수 있기 때문이다.

표 3.3을 보면 분석용, 검증용 자료 모두 일정한 특징을 보이고 있다. 가장 높은 정확도를 보여주는 절단값 0.45를 기준으로 해서 절단값으로 부터 멀어질수록 대체로 정확도는 낮아지는 경향을 보이고, 절단값이 높아질수록 민감도는 낮아지는 경향을, 그리고 특이도는 높아지는 특징을 보이고 있다. 따라서 정확도, 민감도, 특이도가 모두 높은 절단값을 선택하기가 어렵다. 정확도와 민감도의 관점에서는 절단값 0.45가 적절하나 두 자료간에 특이도의 차이가 0.27로서 매우 커서 안정적이지 않다. 따라서 분석용 자료와 검증용 자료의 안정성도 고려하면서 높은 정확도 (분석용 89%, 검증용 82%)를 주는 절단값 0.50을 향후 화학산업의 예측을 위한 절단값으로 선택하였다.

표 4.1에는 화학산업에서와 같은 방법을 적용하여 로지스틱 회귀분석에 의해 각 산업에 대해 선택된 절단값과 선택된 절단값을 기준으로 분류의 정확도, 민감도, 특이도를 분석용 자료와 검증용 자료별로 보여준다.



표 3.3 절단값에 따른 정확도, 민감도, 특이도의 비교  
(목표변수: 화학산업, 모형: 로지스틱 회귀모형)

절단값	분석용 자료			검증용 자료		
	정확도	민감도	특이도	정확도	민감도	특이도
0.00	0.58	1.00	0.00	0.71	1.00	0.00
0.05	0.69	0.96	0.32	0.73	1.00	0.08
0.10	0.69	0.96	0.32	0.73	0.97	0.15
0.15	0.76	0.96	0.47	0.76	0.97	0.23
0.20	0.78	0.96	0.53	0.76	0.97	0.23
0.25	0.78	0.92	0.58	0.82	0.97	0.46
0.30	0.80	0.92	0.63	0.82	0.97	0.46
0.35	0.82	0.92	0.68	0.82	0.94	0.54
0.40	0.89	0.92	0.84	0.82	0.94	0.54
0.45	0.89	0.92	0.84	0.83	0.94	0.57
0.50	0.89	0.92	0.84	0.82	0.88	0.69
0.55	0.89	0.92	0.84	0.78	0.81	0.69
0.60	0.89	0.88	0.89	0.78	0.78	0.77
0.65	0.87	0.85	0.89	0.80	0.75	0.92
0.70	0.82	0.77	0.89	0.80	0.75	0.92
0.75	0.80	0.73	0.89	0.82	0.75	1.00
0.80	0.82	0.73	0.95	0.80	0.72	1.00
0.85	0.78	0.62	1.00	0.78	0.69	1.00
0.90	0.67	0.42	1.00	0.71	0.59	1.00
0.95	0.62	0.35	1.00	0.58	0.41	1.00
1.00	0.42	0.00	1.00	0.29	0.00	1.00

### 3.3. 신경망분석

신경망 (Neural network) 모형은 복잡한 구조를 가진 자료의 예측에 사용되는 비선형모형이다. 신경망 모형들 중에서 가장 널리 사용되는 모형은 다층인식자 (Multilayer perceptron: MLP) 신경망이다. MLP 신경망은 입력층, 은닉마디로 구성된 은닉층, 그리고 출력층으로 구성된 전방향 (Forward) 신경망의 구조를 갖는다.

MLP 신경망모형을 적용시키기에 앞서 은닉층과 은닉마디의 개수를 정해야 한다. 은닉층과 은닉마디의 수가 많아지면 모형이 복잡해지고 추정해야 할 계수의 수가 급격히 증가하여 최적화가 어려워진다. 본 논문에서는 MLP 신경망 모형을 기본으로 하여 최대 2개의 은닉층과 최대 3개의 은닉마디를 갖는 총 12개의 MLP 신경망모형에 대하여 적합시켰다.

표 3.4는 화학산업의 분석용자료와 검증용자료에 대하여 12개의 MLP 신경망모형을 적합시켜 얻은 평균제곱오차 (Mean squared error; MSE)와 오분류율 (Misclassification rate)을 나타낸다. 이 때, 첫 열의 숫자 형식  $h(h_1, h_2)$ 에서  $h$ 는 은닉층의 개수를,  $h_1$ 은 첫 번째 은닉층의 은닉마디 개수를,  $h_2$ 는 두 번째 은닉층의 은닉마디 개수를 나타낸다. 표 3.4에서 MSE와 오분류율이 작으며, MSE와 오분류율에 있어서 분석용 및 검증용 자료의 차이가 작으며, 그리고 추정해야 할 계수의 수가 작은 모형으로서 은닉층의 개수가 2개이고 은닉마디가 각각 1, 2인 2(1,2) MLP 모형을 화학산업의 신경망모형으로 선택하였다.

표 3.5는 각 산업별로 분석용 데이터와 검증용 데이터의 MSE와 오분류율을 기준으로 선택한 MLP 신경망모형의 결과이다. 기계, 건설, 운수창고, 유통산업은 은닉층의 개수가 1개인 모형이 선택되었으며, 나머지 산업들은 은닉층의 개수가 2개인 산업이 선택되었다.

표 3.6은 화학산업의 주가지수 자료에 2(1,2) MLP 신경망모형을 적합시켰을 때 절단값별로 정확도, 민감도 및 특이도를 나타낸다. 앞서 로지스틱 회귀모형에서 절단값을 구한 방법과 같이 분석용 자료와 검증용 자료에서의 안정성을 고려하면서, 정확도, 민감도, 특이도가 높은 절단값으로 0.60을 선택하였다.

표 3.4 은닉층 및 은닉마디의 수에 따른 MSE와 오분류율  
(목표변수: 화학산업, 모형: MLP 신경망모형)

은닉층 (은닉마디)	분석용 자료		검증용 자료	
	MSE	오분류율	MSE	오분류율
1 (1)	0.113	0.111	0.118	0.178
1 (2)	0.099	0.067	0.124	0.178
1 (3)	0.251	0.156	0.144	0.178
2 (1,1)	0.127	0.133	0.116	0.200
2 (1,2)	0.136	0.089	0.116	0.133
2 (1,3)	0.162	0.089	0.117	0.178
2 (2,1)	0.113	0.067	0.123	0.156
2 (2,2)	0.190	0.089	0.111	0.178
2 (2,3)	0.216	0.089	0.135	0.200
2 (3,1)	0.235	0.111	0.141	0.244
2 (3,2)	0.239	0.067	0.118	0.156
2 (3,3)	0.507	0.067	0.134	0.178

표 3.5 목표변수별 MLP 신경망모형의 MSE와 오분류율

목표변수	변수설명	은닉층 (은닉마디)	분석용 자료		검증용 자료	
			MSE	오분류율	MSE	오분류율
y1	제조산업	2(1,1)	0.149	0.148	0.222	0.244
y2	기계산업	1(1)	0.168	0.207	0.244	0.333
y3	건설산업	1(2)	0.258	0.333	0.214	0.333
y4	철강산업	2(1,1)	0.304	0.356	0.226	0.356
y5	음식료산업	2(1,2)	0.248	0.244	0.189	0.289
y6	전기전자산업	2(1,2)	0.213	0.200	0.143	0.222
y7	운수창고산업	1(1)	0.179	0.178	0.174	0.222
y8	금융산업	2(1,2)	0.212	0.222	0.228	0.378
y9	유통산업	1(1)	0.230	0.172	0.172	0.178
y10	서비스산업	2(1,1)	0.193	0.267	0.124	0.200
y11	화학산업	2(1,2)	0.136	0.089	0.116	0.133

표 3.6 절단값에 따른 정확도, 민감도, 특이도의 비교  
(목표변수: 화학산업, 모형: 2(1,2) MLP 신경망모형)

절단값	분석용 자료			검증용 자료		
	정확도	민감도	특이도	정확도	민감도	특이도
0.00	0.58	1.00	0.00	0.71	1.00	0.00
0.05	0.58	1.00	0.00	0.71	1.00	0.00
0.10	0.58	1.00	0.00	0.71	1.00	0.00
0.15	0.58	1.00	0.00	0.71	1.00	0.00
0.20	0.71	0.92	0.42	0.80	1.00	0.31
0.25	0.82	0.88	0.74	0.82	0.91	0.62
0.30	0.84	0.88	0.79	0.82	0.88	0.69
0.35	0.87	0.88	0.84	0.80	0.84	0.69
0.40	0.58	0.58	0.57	0.82	0.84	0.77
0.45	0.87	0.85	0.89	0.84	0.84	0.85
0.50	0.84	0.81	0.89	0.84	0.84	0.85
0.55	0.84	0.81	0.89	0.84	0.81	0.92
0.60	0.87	0.81	0.95	0.84	0.81	0.92
0.65	0.89	0.81	1.00	0.80	0.75	0.92
0.70	0.82	0.69	1.00	0.80	0.72	1.00
0.75	0.82	0.69	1.00	0.78	0.69	1.00
0.80	0.80	0.65	1.00	0.78	0.69	1.00
0.85	0.78	0.62	1.00	0.73	0.63	1.00
0.90	0.71	0.50	1.00	0.69	0.56	1.00
0.95	0.69	0.46	1.00	0.67	0.53	1.00
1.00	0.27	0.00	0.42	0.29	0.00	1.00

표 4.1에는 화학산업에서와 같은 방법을 적용하여 MLP 신경망분석에 의해 각 산업에 대해 선택된 절단값과 선택된 절단값을 기준으로 분류의 정확도, 민감도, 특이도를 분석용 자료와 검증용 자료별로 보여준다.

#### 4. 모형평가

3절에서는 화학산업 주가지수 자료에 대하여 로지스틱 회귀모형과 신경망모형을 적합시켜 절단값에 따른 정확도, 민감도, 특이도를 고려하여 하나의 절단값을 선택하였다. 같은 방법을 적용하여 각 산업에 대한 통계값들이 표 4.1에 보여진다.

표 4.1로부터 분석용자료와 검증용자료의 안정성평가를 해보자. 제조산업 (y1)과 건설산업 (y3)에 대한 로지스틱 회귀모형과 신경망모형에 의한 적합에서, 철강산업 (y4), 음식료산업 (y5), 및 화학산업 (y11)의 로지스틱 회귀모형에 의한 적합에서, 그리고 금융산업 (y8)의 신경망모형의 적합에서 분석용자료와 검증용자료의 특이도가 약 17%p~27%p 정도의 차이가 난다. 또한 금융산업 (y8)의 로지스틱 회귀모형의 적합에서 정확도 및 민감도가 약 20%p 내외의 차이가 나며, 유통산업 (y9)의 신경망모형의 분석에서는 28%p의 민감도가 차이가 난다. 모형의 안정성 측면에서 운수창고산업 (y7)에 대한 로지스틱 회귀모형 및 신경망모형이, 그리고 서비스산업 (y10)과 화학산업 (y11)에 대한 신경망모형이 안정적이라 판단된다.

표 4.1 선택된 절단값에서 각 산업 및 모형별 정확도, 민감도, 특이도의 비교

목표변수	변수설명	모형	절단값	분석용 자료			검증용 자료		
				정확도	민감도	특이도	정확도	민감도	특이도
y1	제조산업	로지스틱	0.45	0.84	0.81	0.88	0.77	0.83	0.64
		신경망	0.50	0.82	0.76	0.88	0.76	0.77	0.71
y2	기계산업	로지스틱	0.20	0.78	0.96	0.57	0.77	0.80	0.71
		신경망	0.40	0.78	0.92	0.62	0.78	0.81	0.71
y3	건설산업	로지스틱	0.45	0.73	0.78	0.68	0.69	0.83	0.44
		신경망	0.60	0.69	0.57	0.82	0.62	0.66	0.56
y4	철강산업	로지스틱	0.55	0.84	0.89	0.76	0.71	0.85	0.50
		신경망	0.60	0.62	0.64	0.59	0.58	0.59	0.56
y5	음식료산업	로지스틱	0.45	0.82	0.86	0.78	0.64	0.70	0.53
		신경망	0.55	0.69	0.59	0.78	0.64	0.60	0.73
y6	전기전자산업	로지스틱	0.55	0.78	0.65	0.88	0.76	0.74	0.78
		신경망	0.50	0.80	0.65	0.92	0.78	0.74	0.83
y7	운수창고산업	로지스틱	0.55	0.82	0.75	0.88	0.80	0.72	0.90
		신경망	0.45	0.78	0.80	0.76	0.78	0.80	0.75
y8	금융산업	로지스틱	0.30	0.82	0.85	0.79	0.66	0.65	0.68
		신경망	0.70	0.87	0.71	1.00	0.62	0.52	0.81
y9	유통산업	로지스틱	0.55	0.75	0.69	0.81	0.77	0.74	0.83
		신경망	0.55	0.69	0.61	0.77	0.82	0.89	0.72
y10	서비스산업	로지스틱	0.60	0.78	0.71	0.83	0.84	0.84	0.86
		신경망	0.70	0.80	0.67	0.92	0.80	0.74	0.93
y11	화학산업	로지스틱	0.50	0.89	0.92	0.84	0.82	0.88	0.69
		신경망	0.60	0.87	0.81	0.95	0.84	0.81	0.92

#### 5. 예측

3절에서 구축한 로지스틱 회귀모형과 신경망모형을 이용하여 2008년 7월부터 2011년 12월까지의 총 42개 월별 자료로 구성된 예측용 자료에 대해 주가상승 혹은 주가하락을 예측해 보기로 하자. 예측용 자

료의 입력 변수들의 값이 주어졌을 때 분석용 자료에 의해 추정된 로지스틱 회귀모형과 신경망모형에 의해 목표변수가 1이 될 확률, 즉, 주가가 상승할 사후 확률을 구할 수 있다. 사후 확률이 표 4.1에 나타나 있는 각 산업별 절단값 보다 높으면 목표변수를 범주 1 (주가 상승)로 분류하고, 사후 확률이 절단값 보다 낮으면 목표변수를 범주 0 (주가 하락)으로 분류하였다. 이렇게 분류한 목표변수 범주와 실제 목표변수 범주를 비교해 얼마나 잘 예측하였는지 예측 정확도를 계산할 수 있다.

표 5.1은 11개의 산업에 대한 예측정확도를 나타낸다. 제조산업 (y1), 건설산업 (y3), 철강산업 (y4), 음식료산업 (y5), 금융산업 (y8), 유통산업 (y9), 서비스산업 (y10), 화학산업 (y11)의 경우 로지스틱 회귀모형 혹은 신경망모형에 의해 최대 약 67%~76%의 정확도를 보여주었으나, 기계산업 (y2), 전기전자산업 (y6), 그리고 운수창고산업 (y7)의 경우에는 최대 약 52%~57%의 정확도를 보여주었다. 특히 운수창고산업 (y7)의 경우 제 4절에서 분석용자료와 검증용자료의 분석결과가 매우 비슷하여 안정적인 결과를 주었던 산업이었지만 실제 예측에서는 정확도가 매우 낮은 편이다. 철강산업 (y4)에서 로지스틱 회귀모형은 신경망모형에 비하여 24%p 더 높은 정확도를 보여준 반면, 건설산업 (y3), 음식료산업 (y5), 그리고 금융산업 (y8)에서는 신경망모형이 로지스틱 회귀모형에 비하여 각각 5%p, 9%p, 10%p 더 높은 정확도를 보여주었다.

표 5.1에서의 예측용 자료에 대한 예측의 정확도는 표 4.1에서의 분석용 및 검증용 자료에 대한 정확도에 비해서 대략 10%p~27%p 까지 떨어짐을 알 수 있다. 특히, 기계산업 (y2), 전기전자산업 (y6), 그리고 운수창고산업 (y7)의 경우는 정확도가 최대 21%p~27%p 하락하였을 뿐만 아니라 정확도는 60%이하이다. 이것은 예측용 자료가 2008년 국제금융위기 이후의 자료로 구성되어 있음을 감안하면 2008년 이전과 달라진 경제환경에 기인한다고 생각된다. 건설산업 (y3)에서 신경망모형을 적용한 경우에 분석용, 검증용, 그리고 예측용자료에 대한 정확도는 각각 0.69, 0.62, 0.67로서 비교적 높으면서 안정적이다.

표 5.1 예측용 자료에 대한 예측의 정확도

목표변수	변수설명	로지스틱 회귀모형	신경망모형
y1	제조산업	0.69	0.69
y2	기계산업	0.52	0.52
y3	건설산업	0.62	0.67
y4	철강산업	0.74	0.50
y5	음식료산업	0.67	0.76
y6	전기전자산업	0.55	0.57
y7	운수창고산업	0.55	0.57
y8	금융산업	0.57	0.67
y9	유통산업	0.64	0.67
y10	서비스산업	0.69	0.71
y11	화학산업	0.69	0.69

## 6. 결론

기존까지의 많은 연구가 KOSPI의 예측을 다루었던 반면, 본 연구에서는 11개 주요 산업별로 구분하여 산업별 주가지수 예측을 수행하였다. 분석에 사용된 자료는 2001년부터 2011년까지의 총 132개의 월별자료를 이용하였다. 현재 해외 경제변수에 의해 영향을 많이 받는 주식시장을 반영하여 우리나라에 가장 많은 영향을 끼치는 미국, 일본, 중국, 유럽의 환율 등 주요 해외 경제지표를 국내 주요 경제지표와 더불어 입력변수로 사용하였다. 로지스틱 회귀분석을 이용하여 각 산업별로 유의한 예측변수를 추출하고 신경망 모형과 더불어 산업별 주가지수의 상승 및 하락을 예측하였다. 로지스틱 회귀모형으로 각 산업별 주가지수에 영향을 미치는 변수를 선택해본 결과, 기계산업, 음식료산업, 금융산업, 서비스산업이

다른 업종에 비하여 해외변수에 더 큰 영향을 받았다. 또한 다우존스와 WTI 국제유가는 대부분의 산업에 영향을 주었다. 본 연구에서는 로지스틱 회귀분석 혹은 신경망분석에 의해 60% 내외의 산업별 주가지수의 예측 정확도를 산출하였다.

본 연구 결과는 종합적인 주가지수를 예측하기 보다는 각 산업별로 구분하여 유의한 변수를 찾아내고 예측을 하면 좀 더 정확한 모형이 될 수 있음을 시사한다. 하지만 경기선행지수, 소비자기대지수, 산업생산지수 등 대부분 한 달에 한번 발표되는 변수로 인하여 주가지수의 월별자료를 이용하였다. 즉, 일별이나 주별 주가지수를 예측하는데 있어서는 한계가 있었다.

## 참고문헌

- 강현철, 한상태, 최종후, 이성건, 김은석, 엄익현, 김미경 (2008). <고객관리 (CRM)를 위한 데이터마이닝방법론>, 자유아카데미, 서울.
- 김삼용, 김진아 (2009). 일반 자기회귀 이분산 모형을 이용한 시계열 자료 분석. <한국데이터정보과학회지>, **20**, 475-483.
- 김유일, 신은경, 홍태호 (2004). 신경망과 SVM을 이용한 주가지수예측. <인터넷상거래연구>, **4**, 221-243.
- 박범조 (2011). 개별 주가에 반영된 시변 무리행동 연구. <한국데이터정보과학회지>, **22**, 423-436.
- 박인찬, 권오진, 김태운 (2011). 시계열 모형을 이용한 주가지수 방향성 예측. <한국데이터정보과학회지>, **20**, 991-998.
- 박주환, 정진혁 (2008). <내일 주가는 상승할까, 하락할까? -SVM을 활용한 KOSPI 예측>, 대신증권 리서치센터 투자전략부.
- 신동근, 정경용 (2011). 웨이블릿 변환과 퍼지 신경망을 이용한 단기 KOSPI 예측. <한국콘텐츠학회논문지>, **11**, 1-7.
- 오경주, 김태운, 정기웅, 김치호 (2011). 선형 및 신경망 자기회귀모형을 이용한 주식시장 불안정지수 개발. <한국데이터정보과학회지>, **22**, 335-351.
- 이은진, 민철홍, 김태선 (2008). 신경 회로망과 통계적 기법을 이용한 종합주가지수 예측 모형의 개발. <전자공학회논문지>, **45**, 95-101.
- 한나영 (2011). <인공신경망을 이용한 KOSPI 예측에 대한 연구>, 석사학위논문, 고려대학교, 서울.
- Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, **55**, 307-319.
- Lee, H. Y. (2008). Combined model of artificial intelligent predictors based on Genetic algorithm for the prediction of KOSPI. *Entrue Journal of Information Technology*, **7**, 33-43.
- SAS Institute Inc. (2000). *Enterprise miner software : Applying data mining techniques*, SAS Institute Inc., Cary, NC.

## Prediction of the industrial stock price index using domestic and foreign economic indices<sup>†</sup>

Iksun Choi<sup>1</sup> · Dongsik Kang<sup>2</sup> · Jungho Lee<sup>3</sup> · Minwoo Kang<sup>4</sup> ·  
Dayoung Song<sup>5</sup> · Seohee Shin<sup>6</sup> · Young Sook Son<sup>7</sup>

<sup>1234567</sup>Department of Statistics, Chonnam National University

Received 30 January 2012, revised 1 March 2012, accepted 6 March 2012

### Abstract

In this paper, we predicted the rise or the fall in eleven major industrial stock price indices unlike existing studies dealing with the prediction of KOSPI that combines all industries. We used as input variables not only domestic economic indices but also foreign economic indices including the U.S.A, Japan, China and Europe that have affected Korean stock market. Numerical analysis through SAS E-miner showed above or below about 60% accuracy using the logistic regression and neural network model.

**Keywords:** Domestic and foreign economic indices, logistic regression model, neural network model, prediction of the industrial stock price index, SAS E-miner.

---

<sup>†</sup> This study was financially supported by Chonnam National University, 2010.

<sup>1</sup> Bachelor of Science, Department of Statistics, Chonnam National University, Buk-Ku, Gwangju 500-757, Korea.

<sup>2</sup> Undergraduate student, Department of Statistics, Chonnam National University, Buk-Ku, Gwangju 500-757, Korea.

<sup>3</sup> Undergraduate student, Department of Statistics, Chonnam National University, Buk-Ku, Gwangju 500-757, Korea.

<sup>4</sup> Undergraduate student, Department of Statistics, Chonnam National University, Buk-Ku, Gwangju 500-757, Korea.

<sup>5</sup> Bachelor of Science, Department of Statistics, Chonnam National University, Buk-Ku, Gwangju 500-757, Korea.

<sup>6</sup> Undergraduate student, Department of Statistics, Chonnam National University, Buk-Ku, Gwangju 500-757, Korea.

<sup>7</sup> Corresponding author: Professor, Department of Statistics, Chonnam National University, Buk-Ku, Gwangju 500-757, Korea.