러프집합을 이용한 주가수익률 예측방법론의 통합모형 개발에 관한 연구

-제조업을 중심으로-

이 훈영* · 박 기남**

Ⅰ. 서 론

한국은 IMF를 맞아 국제사회로부터 기업의 투명성이 문제점으로 지적되어왔다. 유리알처럼 투명한 기업이어야 국제경쟁력을 갖출 수 있고 그 기업에 대한 위험과 기대가 제대로 반영되어 투자자나 채권자들이 올바른 의사결정을 내릴 수 있다. 또한 이렇게 함으로써 시장의 왜곡을 방지할 수 있다. 그러나 지금까지 국내 기업들은 심지어 대기업까지도 회계분식과 회계처리방법의 잦은 변경, 자산재평가 등 갖가지 방법들을 동원하여 기업의 목적에 맞도록 회계정보를 고쳐온 것으로 드러나고 있다.

그러나 이러한 재무제표의 신뢰성에는 한계점이 있다 하더라도 재무제표 정보를 자본시장에 활용하려는 노력이 계속되어 왔고 회계변수 특히 이익정보가 주가수익률을 설명한다는 사실이 밝혀지면서 재무제표분석을 통하여 내재적 가치를 분석하고 이러한 정보를 투자에 적극 활용하려는 움직임 일어나게 되었다.(정혜영,1995)

재무제표정보를 이용한 이익예측연구는 국내외에서 활발하게 이루어져 회계변수의 정보내용과 예측능력에 관한 실증적 증거를 제공했다(Ou,1984; Ou and Penman,1989;Holthausen and Larcker,1992). 따라서 재무제표를 활용한 투자방법에 획기적인 전환을 위해서는 적절하게 회계정보시스템의 예측모형들을 적극적으로 활용하는 것이 도움이 될 수 있으며(서영호,1998;이훈영,박기남,정혜영). 본 연구에서는 지금까지 이익예측연구에서 사용되어 온 단일 모형들을 결합하는 통합모형을 제시하고자 한다.

이를 위해 전통적으로 통계적 방법인 Logit 분석을 주로 이용하였으며 전통 재무회계에서는 현재까지도 이 방법만을 고수하고 있다. 그러나 1980년대 후반부터 AI(Artificial Intelligence) 분야에서 인공신경망(Artificial Neural Network), 사례기반추론, ID3, CART, 유전자 알고리즘, MARS 등과 같은 방법론에 대한 연구가 이루어지면서 다양한 관점의 방법론들이 대두되었고 이러한 방법론들을 전통적인 회계문제나 재무문제에 적용시킴으로써 의사결정의 품질을 한차원 더 높이기 위한 연구들이 계속되어 왔다.

특히 이러한 연구들은 경영분야의 다양한 영역에서 방법론간의 성과비교 연구의 방식으로 이루어져 각 방법론들의 특징이 어떠한 영역에 보다 적합한지를 고찰하려는 목적의 연구였다. 그러나 개별 방법론들은 모두 독특한 특성을 소유하고 있어서 고유한 장점과 단점이 존재하기 때문에 방법론들을 통합함으로써 방법론간의 장점을 최대화하고 단점을 최소화하는 통합모형의 개발이 현재까지 큰 이슈가 되어오고 있다(이건창,한인구,김명종;1996;한인구,신경식,1999).

이에 따라 성과의 개선을 가져올 수 있는 통합모형의 개발을 위하여 다양한 연구들이 이루어졌으나 지금까지의 통합모형에 대한 연구는 각각의 단일모형의 단점을 극복하고자 단일 모

* 경희대학교 경영학부 부교수
** 경희대학교 경영학부 박사과정

형들을 결합한 통합모형의 구성에 관심이 집중되어왔다.

통합모형은 단일모형이 가지고 있는 장점과 특성을 수용하여 단일모형보다 안정성과 성능의 향상을 얻고자 하는 것이 목적이다. 그러나 통합모형은 모형들간의 결합에 대한 논리적인 기준의 마련이 쉽지 않고 각 모형들간의 결과 값이 상충되는 경우에 이를 해결하는 규칙이 필요하다. 그러나 본 연구의 관점은 개별 모형들의 장점을 살린 결과값을 산출하고 각 모형들이 산출한 결과 값을 바탕으로 모형을 통합한다. 다시 말해서 모형들간의 결과값이 상충되는 문제를 해결하면서 자연스럽게 여러 모형의 결과를 통합하는 방법론으로 러프집합을 사용하는 것이다.

지금까지 주가수익률의 예측에 적용된 대표적 통계적 모형인 로짓 모형을 AI 방법론의 모형들의 성과를 비교하고 한걸음 더 나아가서 러프집합을 이용한 통합방법론과 개별 방법론간의 성과를 비교하여 본 연구에서 제시하는 통합방법론의 우수성을 입증하고자 한다.

본 연구는 다음과 같이 진행된다. 서론에 이어 2 장에는 연구설계, 실증모형, 사용된 데이터셋, 변수 측정에 관하여 설명하고 실증결과는 3 장에서 논의되며, 마지막 장에서는 결론과 향후 연구과제가 제시될 것이다

Ⅱ. 선행연구

자본시장이 효율적이라면 주가는 기업에 대한 모든 정보를 포함하고 있어야 한다. Beaver, Lambert 와 Morse(1980), Beaver, Lambert 와 Ryan(1987), Collins, Kothari 와 Rayburn(1987) 등에서는 주가가 사후적으로 공시되는 회계이익수치에 선행하는 것으로 보았다.

즉 주가는 사후적으로 공시되는 이익에 관한 정보를 사전에 반영하고 이익은 주가에 영향을 미치는 사건을 시차를 가지고 반영한다는 것이다. Beaver, Lambert 와 Morse(1980)등과는 달리 OP(1989)는 미래이익을 예측하는 주가정보가 재무제표내에 포함되어 있다는 견해를 보인다. 즉, 회계이익은 주가에 반영된 정보를 반영하는 데 시차를 가지고 있지만 재무제표에 내재된 다양한 정보는 주가가 지니는 정보이상을 포함한다는 것이다.

OP(1989)는 미래이익의 예측력이 기대되는 68 개의 재무비율을 선정하고 logit 분석을 통해 차기 이익에 대한 예측력이 있는 변수를 선택한 후 미래이익의 증가확률(Pr)을 계산하였다. 미래이익예측모형의 추정기간은 1965-72 년 1973-77 년의 두 기간이며 투자전략의 수익성은 1973-77 년, 1978-83 년의 두 기간에 걸쳐 검증되었다. Pr 값에 의거한 미래이익변화 방향의 예측정확도는 60-67%였으며 Pr 값에 따라 순투자가 zero 인 매수와 매도투자전략을 취한 결과 2 년간의 시장조정수익률이 평균 14.5%로 나타났다. 이 수익률은 기업의 위험을 나타내는 다양한 변수들로도 설명이 되지 않았으므로 재무제표분석을 통한 주식수익률의 예측이 가능하다는 결론을 내리고 있다.

Holthausen and Larcker(1991)는 미래 주식 수익률의 증감을 직접 예측하는 방법으로 회계정보로부터 이익을 얻을 수 있는 거래전략이 가능함을 보였다. 이 연구는 OP(1989)가 비정상수익률을 직접 예측하지 않고 미래이익의 예측을 통해 수익률을 간접평가하고 있는 문제를 제기하고 이러한 문제를 감소시키기 위해 시장조정수익률, 규모조정수익률과 같은 세 가지 대체적인 측정치를 사용하였다. 비정상수익률 예측에 따른 예측기간의 Pr 수치를 10 등분하여 포트폴리오를 구성하고 검증기간의 비정상수익률 예측능력을 평가하였다.

각 비정상수익률 예측모형의 수익성은 12 개월간의 보유 수익률로 나타나는데 1978-88 년의 기간동안 평균 연간비정상수익률은 4.3%0-7.9%로

나타났다. 회귀식을 이용하여 이상현상을 통제한 추가분석에서도 모든 수익률 측정치에서 비정상 수익률이 나타나고 있어 재무제표를 이용하여 주가의 움직임을 예측할 수 있다는 OP(1989)의 주장을 지지하고 있다.

국내자료를 이용한 정혜영(1991)의 연구에서는 30 여 개의 구체적인 비율을 대상으로 단순 logit 분석을 행하여 7 개의 재무비율을 선정하고 선정된 비율을 이용하여 Pr 을 추정하고 이에 따라 투자전략을 수립하여 누적비정상수익률을 분석하였다.

추정기간은 1981 년-1984 년의 경우 다음 해인 1985 년 자료를 이용하여 투자전략을 세우고 1981 년-1985 년을 추정기간으로 하는 경우 1986 년 자료를 이용하는 방식으로 1985 년에서 1988 년까지의 4 년간의 검증기간에서 투자전략에 따른 1 년간 보유수익률을 보고하고 있다. 이 연구에서는 1980 년 1 월 1 일에 존재하는 143 개의 기업을 대상으로 Pr 값을 이용한 투자전략을 수행한 결과 23.2%, 29.5%, 44.8%의 누적비정상수익률을 보이고 있다.

전통적인 로짓분석과 AI 방법론간의 이익예측 성과를 비교분석한 연구로는 Charitou and Charalambous(1996)와 이훈영,박기남,정혜영(1999)의 연구가 있고 두 연구 모두 Ou and Penman 의 연구를 기반으로 하였다. 또한 Holthausen and Larcker(1991)의 연구처럼 주가수익률을 곧바로 예측할 때 AI 기법을 사용한 연구로는 이훈영,박기남,임정혁(1999)의 연구가 유일하다.

Ⅲ. 연구 방법론

본 연구에서는 네 개의 개별모형과 세 개의 통합모형을 1986-1996 년까지 주가수익률의 변화를 예측하기 위해 사용하였는데 로짓모형, 인공신경망, MARS, QDA 가 그것이다. 실증모형에 대한 개괄적인 설명은 다음과 같다.

3.1 로짓분석

다변량 로짓분석은 Ohlson(1980)이 도산예측연구에 도입한 이후 각종 회계연구에서 많이 활용되고 있는 통계적 기법이다. 이익예측연구에서는 Ou(1984,1990), Ou and Penman (1989), Holthausen and Larcker(1992)의 연구에서 그리고 국내에서는 정혜영(1991)의 연구에서 로짓분석모형이 이익변화를 예측하는데 사용된 바 있다.

로짓모형은 예측뿐만 아니라 독립변수의 계수를 추정하고 검증하는데도 이용되고 있다. 로짓모형에서 주가수익률 it+1>0 의 확률을 구하게 되는데 그것을 수식으로 나타내면 다음과 같다.

Prob(주가수익률 it+1 > 0) = exp(z)/1+exp(z)

여기서 e 는 자연로그 값을 의미하고, Z 는 각 다음 수식과 같은 선형결합의 형태를 갖게 된다.

$$Z_K = W_0 + \sum_{i=0}^{n} W_I X_I$$

K=1,2,3,.....,m

이때 로지스틱 회귀선은 S 자 형태의 곡선형태를 가지며, 0 과 1 사이의 확률값을 갖게 된다. 또한 본 연구에서는 모형의 파라미터를 추정하는 방법으로 가장 많이 이용되는 최대우도법(maximum likelihood method; MLE)을 사용하였다. 따라서 예측되어질 종속변수는 t+1 기의 주가수익률의 변화이고 로짓모형의 결과는 t+1 기에 양의(+) 주가수익률변화의 가능성을 추정한다. 또한 추정된 주가수익률의 변화확률은 0 에서 1 까지로 나타나며 확률값이 0.5 이상이면 주가수익률이 증가될 것이라는 예측으로 보고 확률값이 0.5 미만일 때 주가수익률이 감소할 것이라는 예측으로 보았다.

3.2 인공신경망 모형

본 연구에서 로짓모형의 정확성과 안정성을 살펴보기 위해 인공신경망을 사용하고 그 결과를 비교하였다. 인공신경망은 인간두뇌의 휴리스틱(heuriseics)한 문제해결 방법을 모형화한 것으

로서 그 학습능력과 추론능력이 매우 뛰어난 것으로 알려져 있다. 인공신경망의 구조는 여러가지 형태가 있으나 가장 일반적으로 많이 쓰이는 형태는 다층전향구조(multilayerd feedforward) 인공신경망이다.(1993. 이건창)

이는 입력층(input layer),은닉층(hidden layer), 출력층(output layer)의 삼층구조를 이루고 각 층마다 다수의 뉴런 또는 노드(node) 즉 처리단위(processing unit)를 소유하고 있다.서로 다른 층에 존재하는 처리단위는 서로 연결되어 있으며 그러한 연결의 강도는 연결가중치로 계산된다. 결국 인공신경망은 주어진 문제를 해결하기 위한 지식을 이러한 다수의 연결가중치로 분산표현하고 있으며 따라서 이는 기존의 학습방법에서는 찾아볼 수 없는 강력한 학습성(learnability)과 견고성(robustness)을 갖고 있다. 인공신경망은 생물학적 뇌에 있어서 뉴런(Neuron)에 해당하는 처리요소(processing elements;PE)로 구성된다. 처리 요소는 입력, 순입력, 출력, 가중치, 전이함수 등 5개의 기본요소로 이루어진다.

3.3 MARS

Frideman(1991)이 개발한 MARS(Multivariate

$$f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{k=1}^{Lm} h_{lm}(x_{v(l,m)}) \qquad [수식1]$$

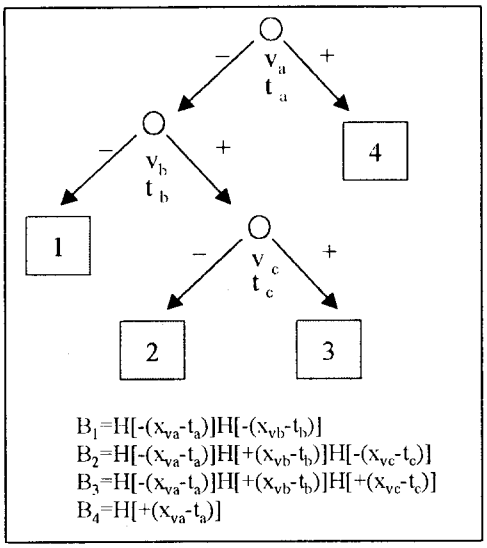
Adaptive Regression Splines)는 적응력이 있는 비모수적 회귀분석으로써 그 형태가 회귀분석의 모형과 비슷하다.

여기서 x1,x2.....xp 는 예측변수이고, v(*l*,*m*)은 m번째 적(products)의 *l* 번째 항(term)에서 사용된 예측변수의 인덱스를 의미한다.다. 이때 기본 함수 h_{lm} 은 쌍으로 정의된다.

$$h_{l,m}(x) = [x - t_{lm}]+$$
$$h_{l,m+1}(x) = [t_{lm} - x]+ \qquad [수식 2]$$

m 이 어떤 정수일 때, knot 값 t_{lm} 은 Xv(*l*,*m*)의 고유한 값 중 하나이다. 수식 1 은 수식 2 형태의 낱개의 선형기본함수들의 적(products)을 합한 것이라는 것을 알 수 있다. 전통적인 표기법으로 Hj(x)로 표현되는 모형은 상수 기본 함수로 출발하여 후보 기본 함수들의 클래스가 모두 [x-t]+형태가 되도록 정의한다. 그 후 매 단계마다 적(product) Hj(x)hlm(x)와 Hj(x)hlm+1(x)으로 정의된 두개의 항(term)이 모형에 추가되어야 하는지 고려한다. 만약 Hj(x)hlm(x)와 Hj(x)hlm+1(x)이 제곱된 잔차의 합(residual sum of square) 을 가장 크게 줄이면 이것은 모형에 추가된다. 이러한 forward stepwise 방식으로 모형을 구성해나가시 최대 모형크기에 도달하면 멈추게 된다.

<그림 1> MARS 의 기본함수 알고리즘



마지막 backward prunning 절차가 모형에 적용되며, 가장 덜 중요한 항을 한번에 하나씩 줄인다. 일련의 단계를 거치면서 가장 적합한 모형이 선택되는데, 이것은 GCV 에 의해서 측정된다.

J class 문제의 정준함수(J-1=K)는 형태로 표현할 수 있는 데 hm(X)는 Basis Function 이라고 한다. MARS 에서는 최적의 입력변수와 그 변수와 관련된 knot 로 이루어진 식과 interactive 한 변수결합(단, degree 가 2 인 경우)으로 이루어진 식으로 함수식이 구성된다.

3.4 BRUTO(adaptive additive modelling)

BRUTO 는 Adaptive additive 회귀분석으로서, MARS 보다 더 엄격하다. 또한 BRUTO 는

smoother fashion으로 예측변수들을 다루기 때문에 예측변수가 많은 상황에서 효과적이다. 모형을 수식으로 표현하면 다음과 같다.

$$\sum_{i=1}^{N}(y_i - \sum_{j=1}^{p}\beta_j^T h_j(x_{ij}))^2 + \sum_{j=1}^{p}\lambda_j \chi_j^T \Omega_j \beta_j \quad [수식3]$$

다행히 식 3은 단순화 될 수 있고 λ_j는 단순한 일련의 추정식을 통하여 표현하는 것이 가능하다.

$$f_j = s_j(\lambda_j)\{y_i - \sum_{k \neq j} f_k(x_{ik})\} \quad [수식4]$$

여기서 Sj는 j 번째 예측변수에 대해 smoothing하기 위한 smoothing spline operator이다. 이러한 수식들은 자연스럽게 backfitting으로 알려져 있는 iterative Gauss-Seidel 알고리즘으로 된다(Hastie,1993).

이때 모형에 추가해야 할 항(term)과 선택된 항들의 λ의 값을 결정해야 하고 복수의 응답변수를 수용할 수 있도록 전체절차를 수정하는 과정이 필요하다. 예를 들어서 p 개의 변수가 주어져서 p 개의 항으로 된 additive 모형이라면 이때 파라미터 λ를 smoothing하는 모든 p에 대해서 GCV를 이용하여 최적화된 모형을 탐색해 내는 방법이다.

3.5 러프집합을 이용한 분류(Classification)

현실세계에서 주로 다루어지는 데이터의 형태는 2 차원 데이블(2-Dimemsional Table)형태의 데이터이다. 2 차원 테이블 형태의 데이터는 통계적 개념에의 독립변수 와 종속변수 그리고 관찰값들로 이루어진 구조를 가지고 있다. 러프집합에서는 이러한 데이터의 형태를 Decision System라고 새롭게 표현되는데 이에 대한 정의는 다음과 같다.

DS=(U, A∪C), A ∩C= φ

여기서 U는 관찰값들의 전체집합, A는 condition attribute 집합, C는 decision attribute 집합이라 한다. 이때 임이의 a∈A 에 대해서 a:U→V_a가 주어질 때 V_a를 a attribute에 대한 value 집합이라 하고 마찬가지로 c∈C에 대해서 c:U→Vc가 주어질 때 Vc를 c attribute에 대한 value 집합이라 정의한다. 그러면 분류 작업을 위해서 Decision System으로부터 IF-THEN 형태의 Rule의 집합을 생성하는데 이때 임이의 x∈U의 decision attribute가 c일 경우 이에 대한 Rule은 다음과 같이 정의된다.

RUL(A,x,c)={$F_B(x)$→(c=c(x)):B∈RED(A,x,c)}

$F_B(x)$=$\Pi_{b\in B}$(b=b(X)),B⊂A,

A는 U의 전체 condition attribute 집합

RED(A,x,c)는 x를 다른 원소와 구별하기 위한 minimal attribute 집합이다. 위의 공식에 의해서 x∈U에 대한 Rule이 생성이 되었는데 모든 U의 원소에 대한 Rule 집합으로 일반화 시키면 다음과 같다.

RUL(A,c)= $\cup_{x\in U}$ RUL(A,x,c)

러프집합을 이용한 분류작업은 Tranning Data Sample에서 생성된 Rule들의 집합에서 Test Data Sample의 각각의 관찰값에 대응(match)되는 Rule을 선택하는 작업이다. 러프집합에서 분류를 하는데 있어서 사용하는 알고리즘은 특별하게 정해져있지는 않다. 따라서 본 연구에서 사용한 알고리즘은 러프집합 분석도구인 *Rosetta™*에서 제공하는 Standard Voter 알고리즘을 사용하였다. 모형 결합문제에 있어서 가장 어려운 점은 각 모형들에 의해서 나온 결과들이 서로 상충되었을 때 이를 적절하게 해결하는 문제인데 이를 해결하기 위해서는 Standard Voter 알고리즘의 분류절차는 다음과 같다.

STEP 1 : 새로운 관찰값 y∈T의 $F_b(y)$ → (c,c(y))가 v∈RUL(A,c)인 v가 존재하면 y∈T는 v에 의해 분류된다.

STEP2: 관찰값 y∈T의 $F_b(y)$ → (c,c(y))가 v∈RUL(A,c)인 v를 발견할 수 없다면 y∈T는 오분

류(misclassification)로 처리한다.

STEP 3 : 관찰값 y∈T 의 $F_b(y) \rightarrow (c,c(y))$가 복수개 존재할 경우 v∈RUL(A,c)인 v 을 선정할 때 certainty(x, β)가 높은 v 를 선택한다.

Ⅳ. 실증분석

4.1 데이터의 구성

표본기업은 KIS-FAS 데이터베이스로부터 선택되었다. 1982-1996 년 기간동안 후보변수들을 계산하기 위해 다음과 같은 조건을 만족하는 회사들을 표본으로 선정하였다.

(1)12 월이 결산인 제조업체이고 (2)자료를 구성하는데 필요한 기간(3 년)과 모형추정을 위해 필요한 기간(3 년)동안 재무제표자료가 누락 없이 KIS-FAS 데이터베이스 내에 기록이 되어있고 (4) KIS-SMAT 에 주가수익률자료가 등록되어 있는 기업을 대상으로 <표 1>과 같이 데이터 셋을 구축하였다.

<표 1> 데이터 셋의 구성

데이터셋	F86	F87	F88	F89	F90	F91	F92	F93	F94	F95	F96
추정셋의 기간	1983-1985	1984-1986	1985-1987	1986-1988	1987-1989	1988-1990	1989-1991	1990-1992	1991-1993	1992-1994	1993-1995
총 표본 수	585	561	600	768	948	1186	1324	1388	1390	1398	1444
추정셋의 표본수	447	435	386	478	630	822	972	1034	1070	1026	1046
검증년도	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
검증셋의 표본수	138	126	214	290	318	364	352	354	320	372	398

1982 년부터 1996 년까지의 기간동안 F86 에서 F96 까지 총 11 개의 데이터셀으로 구성되어있고 각각의 연도(t)에 직전년도 3 개 년(t-1,t-2,t-3)의 데이터를 이용하여 모형을 추정한 후 다음해(t+1)의 예측력을 검증한다. 구체적으로 예측모형의 계수추정은 검증대상기간 직전의 3 개년 자료를 통합하여(cross-sectional and time-seres pooled)이용된다.

4.2 변수 측정

예측되어질 종속변수는 회사 i 의 미래 주가수익률의 변화이다. 본 연구에서는 주가수익률을 산출할 때 t 년도의 개별기업의 주가수익률과 t 년도의 시장조정 주가수익률을 비교하여 종속변수를 구성하였다. 본 연구에서 사용될 시장조정 수익률을 다음과 같이 정의한다

$$\text{시장조정수익률} = MAR_{im} = \prod_{t=1}^{m}(1+R_{it}) - \prod_{t=1}^{m}(1+R_{mt})$$

여기서 R_{it} 는 t 월의 기업 i 의 주식 수익률

R_{mt} 는 t 월의 단순평균 주가지수수익률

여기서, 주가수익률의 변화를 나타내는 더미변수로 Dit+1 를 정의하고 이것을 다음과 같은 기준으로 그 값을 0 과 1 의 이진수로 표기한다.

$D_{it+1} = 1$, if 주가수익률$_{it+1} > 0$

$D_{it+1} = 0$, if 주가수익률$_{it+1} < 0$

연속치로 예측하는 모형은 모형설정상의 오류가 있으면 추정치가 편의를 갖게 되고 관련 통계량이 과대하게 나타나는 문제점이 있다 (Freeman, Ohlson,Penman, 1982).

4.3 변수선정

예측변수는 Ou(1984,1990), Ou and Penman (1989)에 의해 조사된 68 개의 변수 목록으로부터 시작하여 정혜영(1991)의 연구와 정혜영,이현, 장상기(1995)의 연구에서 통계적으로 유의한 것으로 나타난 변수들과 그 후 재무제표를 활용한 주가수익률 예측이나 내재적 가치에 관련된 문헌연구에서 통계적으로 유의한 변수를 후보변수 (Cadidate Variable)로 포함시켜 최종 58 개의 변수들을 선정하였다.

<표 2> 본 연구에서 선정된 총 변수 목록

변수	변수의 내용	변수	변수의 내용
x1	당기순이익/자기자본	x2	경상이익/자기자본
x3	영업이익/자기자본	x4	(영업이익+감가상각비)/자기자본
x5	이익잉여금/자기자본	x6	(당기순이익-배당금)/자기자본
x7	총부채/자기자본	x8	고정부채/자기자본
x9	자기자본	x10	영업이익/매출액
x11	장기순이익/매출액	x12	법인세비용차감전순이익/매출액
x13	매출 총이익/매출액	x14	(경상이익+감가상각비)/매출액
x15	(영업이익+감가상각비)/매출액	x16	연구개발비/매출액
x17	유동자산/매출액	x18	감가상각비/매출액
x19	(판매관리비-감가상각비)/매출액	x20	고정자산/매출액
x21	(유형자산-유동부채)/매출액	x22	광고선전비/매출액
x23	매출액	x24	매출액/총자산
x25	(경상이익+감가상각비)/총자산	x26	(영업이익+감가상각비)/총자산
x27	연구개발비/총자산	x28	기계장치/총자산
x29	총부채/총자산	x30	재고자산/총자산
x31	현금과 예금/총자산	x32	이익잉여금/총자산
x33	당기순이익/총자산	x34	(기계장치+재고자산+투자자산)/총자산
x35	유형고정자산/총자산	x36	총자산
x37	매출원가/재고자산	x38	매출액/재고자산
x39	재고자산	x40	자기자본/고정자산
x41	감가상각비/유형고정자산	x42	매출액/현금과 예금
x43	매출액/매출채권	x44	배당금/(경상이익+감가상각비)
x45	당기순이익/(경상이익+감가상각비)	x46	(경상이익+감가상각비)/(기계장치+재고자산+배당금)
x47	배당금/당기순이익	x48	기계장치/당기순이익
x49	감가상각비/기계장치	x50	(법인세비용차감전순이익+이자비용)/이자비용
x51	유동자산/유동부채	x52	당좌자산/유동부채
x53	매출액/매출채권	x54	연구개발비
x55	광고선전비	x56	(당기순이익+감가상각비)/총부채
x57	고정부채	x58	이익잉여금

다음 단계로 본 연구에서는 문헌 연구를 통하여 선정된 58 개 변수 각각의 통계적인 유의성을 살펴보고자 단변량 분석인 T-test 분석과 상관분석을 수행하였으며 로짓분석, MARS, BRUTO 모형에는 Stepwise 로짓분석에서 선정된 변수들을 입력변수로 사용하였고 인공신경망의 경우에는 유전자알고리즘을 이용한 변수선정 방법을 이용하였다. 최종적으로 각 방법론에서 사용된 변수들이 <표 3>과 <표 4>에 요약되어 있다.

<표 3> 인공신경망에서 사용된 최종 변수목록

셀 명	기간	변수명
FOR86	1983-1985	X24 X25 X29 X33 X36 X39
FOR87	1984-1986	X10 X14 X15 X22 X24, X30
FOR88	1985-1987	X10 X12 X14 X15 X22 X32
FOR89	1986-1988	X10 X11 X12 X13 X14 X15 X22 X30 X41
FOR90	1987-1989	X10 X13 X14 X15 X19 X22
FOR91	1988-1990	X13 X18 X19 X24 X49
FOR92	1989-1991	X9 X13 X18 X20 X24 X37
FOR93	1990-1992	X20 X24 X25 X37 X38 X45
FOR94	1991-1993	X20 X24 X25 X32 X37
FOR95	1992-1994	X20 X24 X38 X41 X45
FOR96	1993-1995	X3 X4 X27 X29 X30 X32 X39

<표 4> 로짓분석, MARS, BRUTO 에서 사용된 최종 변수목록

셀 명	기간	변수명
FOR86	1983-1985	X24 X28 X36 X39 X41 X51
FOR87	1984-1986	X10 X15 X20 X21 X24 X28 X31 X36 X37 X40 X55

92	0.4801	0.5028	0.5426	0.5852	0.4801	0.5852	0.5681
93	0.4915	0.5141	0.5480	0.5451	0.5225	0.5480	0.5621
94	0.5343	0.5250	0.5875	0.5187	0.5343	0.5406	0.5500
95	0.5483	0.5594	0.5752	0.5645	0.5483	0.5698	0.5698
96	0.5477	0.5276	0.6683	0.5678	0.5577	0.6683	0.6658
평균	0.5142	0.5441	0.5636	0.5517	0.5215	0.5780	0.5863

<표 7> 전체 셋에서 개별모형과 통합모형의 예측정확도

year	LOGIT	NN	MARS	BRUTO	LN통합	MB통합	LNMB통합
86	0.5852	0.5467	0.6058	0.5870	0.5852	0.6010	0.6195
87	0.5639	0.5500	0.5267	0.5533	0.5559	0.5839	0.5932
88	0.5829	0.5730	0.5875	0.5937	0.5805	0.5987	0.6102
89	0.5783	0.5787	0.5230	0.5924	0.6076	0.6066	0.6176
90	0.5160	0.5189	0.5394	0.5168	0.5247	0.5467	0.5522
91	0.5363	0.6088	0.5755	0.5500	0.5295	0.5743	0.5798
92	0.5204	0.5214	0.5496	0.5720	0.5204	0.5750	0.5659
93	0.5189	0.5249	0.5449	0.5481	0.5359	0.5496	0.5493
94	0.5363	0.5073	0.5847	0.5320	0.5363	0.5605	0.5619
95	0.5514	0.5368	0.5641	0.5615	0.5537	0.5446	0.5588
96	0.5573	0.5142	0.6099	0.5573	0.5618	0.7065	0.6565
평균	0.5497	0.5437	0.5646	0.5604	0.5538	0.5861	0.5895

마지막으로 전체 데이터 셋에서 예측정확도를 분석해보면 <표 7>과 같다. 전체 데이터 셋에서의 모형별로는 로짓모형 54.9%, 인공신경망의 경우 54.4%, MARS는 56.5%, BRUTO의 경우에는 56.0%로 나타났다. 통합모형에서는 LN통합모형의 경우 평균 55.4%의 예측 정확도를 보였고 MB통합모형에서는 58.6%로 나타났으며 LNMB통합모형의 경우에는 평균 59.0%의 예측 정확도를 나타냈다.

5.2 개별 모형들간의 예측정확도 차이분석

본 연구에서 사용한 개별 방법론들의 예측정확도간의 차이를 비모수적 기법인 부호순위 검증을 통하여 살펴보았다. 먼저 추정 셋에서의 개별모형들간의 차이를 분석하여 요약한 것이 <표 8>이다. 표에서 모형1과 모형2는 서로간에 비교될 모형들을 의미하며 평균은 두 모형의 차이의 평균을 의미한다. 단일 모형에서는 로짓모형과 신경망모형의 예측정확도가 통계적으로 유의한 차이를 보이는 것으로 나타났고 통합모형과의 비교에서는 인공신경망과 LN통합모형이 유의한 차이를 보였다. LNMB통합모형의 경우 로짓모형을 제외한 모든 모형과의 비교에서 유의한 차이를 보이고 있다.

<표 8> 추정 셋에서 예측정확도 비교

모형1	모형2	평균	부호순위	p
NN	LOGIT	−0.0418	−23	0.0420
LN통합	LOGIT	0.0008	10.5	0.2422
LN통합	NN	0.0427	23	0.0420
BRUTO	MARS	0.0034	3	0.8311
MB통합	MARS	0.0286	22	0.0537
MB통합	BRUTO	0.0252	18	0.1230
LNMB	LOGIT	0.0051	8	0.5195
LNMB	NN	0.0470	25	0.0244
LNMB	LN통합	0.0042	6.5	0.5566
LNMB	MARS	0.0246	28	0.0098
LNMB	BRUTO	0.0213	23	0.0420
LNMB	MB통합	−0.0040	2	0.8984

<표 9> 검증 셋에서 예측정확도 비교

모형1	모형2	mean	부호순위	p
NN	LOGIT	0.0299	28	0.0098
LN통합	LOGIT	0.0073	8.5	0.4180
LN통합	NN	0.0227	21	0.0674

BRUTO	MARS	−0.0119	−9	0.4648
LOGIT	MARS	−0.0494	−30	0.0049
LOGIT	BRUTO	−0.0375	−32	0.0020
NN	BRUTO	−0.0375	−32	0.0020
MB통합	MARS	0.01192	9	0.4648
MB통합	BRUTO	0.0262	30	0.0049
LNMB	LOGIT	0.0721	33	0.0010
LNMB	NN	0.0421	33	0.0010
LNMB	LN통합	0.0647	33	0.0010
LNMB	MARS	0.0226	14	0.2402
LNMB	BRUTO	0.0345	29	0.0068
LNMB	MB통합	0.0082	14.5	0.0977

<표 10> 전체 셋에서 예측정확도 비교

모형1	모형2	mean	부호 순위	p
LOGIT	NN	0.0059	11.5	0.7181
LOGIT	LN통합	0.0041	32	0.2097
NN	LN통합	0.0100	26.5	0.4022
MARS	BRUTO	0.0042	15.5	0.6261
NN	MARS	−0.0209	−61.5	0.0428
MARS	MB통합	0.0214	55.5	0.0703
BRUTO	MB통합	0.0257	95.5	0.0006
LOGIT	LNMB	0.0386	93.5	0.0008
NN	LNMB	0.0445	108.5	0.0001
LN통합	LNMB	0.0345	86.5	0.0009
MARS	LNMB	0.0236	85.5	0.0029
BRUTO	LNMB	0.0279	102.5	0.0002
MB통합	LNMB	0.0021	38	0.1650

검정 셋에서의 개별모형들간의 차이를 분석하면 <표 9>와 같이 요약된다. BRUTO 모형과 MB 통합모형간의 차이가 유의했으며 LNMB 통합모형의 경우 MARS 모형을 제외한 모든 모형과의 비교에서 통계적으로 유의한 차이를 보이고 있다. 전체 셋에서의 개별 모형들간의 차이를 분석한 것이 <표 10>이다. 통합모형과의 비교에서 MB 통합모형과 LNMB 통합모형에서 모든 개별모형보다 통계적으로 유의하게 개선된 결과를 보였다. LNMB 통합모형의 경우 MB 통합모형을 제외한 모든 통합모형과도 통계적으로 유의한 차이를 보이고 있다.

VI. 결 론

본 논문은 회계학연구가 회계정보를 이용하여 주가수익률을 어떻게 예측할 수 있는가에 관한 관심을 가지지 않았다는 점에 주목하여 회계정보를 이용하여 주가수익률을 평가해 보고자 하였다. 경제이론의 측면에서 이익은 신호의 역할을 함으로써 자본시장에서 최적의 자원배분을 하게 한다. 재무분석가 들은 더 나은 주가수익률을 기대하며 증권 및 채권투자를 수행한다. 즉, 주가수익률의 변화를 예측하는데 사용될 수 있는 회계정보를 찾아내고 이러한 회계정보를 이용하여 투자한다고 가정했을 때 기존의 로짓모형과 인공신경망모형에 더하여 MARS 모형과 BRUTO 모형을 소개하고 이들 개별모형을 러프집합을 활용하여 통합하였을 때 통계적으로 유의한 개선효과를 얻을 수 있다는 것을 보임으로써 실제 주식시장에서 적용가능성을 보였다.

참 고 문 헌[1]

정혜영, "재무제표정보와 주가수익률 예측", 회계학연구, 제 12 권, 7 월, 1991, pp.31-56

정혜영, 이현, 장상기, "재무제표분석과 기업의 내재적 가격결정에 의한 투자전략", 회계학연구, 제 20 권, 제 1 호, 3 월, 1995, pp.101-130.

Blume,M. and R. Stambaugh, "Bias in Computed Returns: An Application to the Size Effect, Journal of Financial Economics", October, 1983, pp.557-581.

Charitou, A. and C. Charalambous, "The Prediction of Earnings Using Financial Statement Information: Empirical Evidence With Logit Models and Artificial Neural Networks" Intelligenct Systems in Accounting, Finance and Management, VOL.5, 1996, pp.199-215.

Holthausen, R. and D. Larcker, "The prediction of stock returns using financial statement information',Journal of Accounting and Economics, 15, 1992, pp.373-411.

[1] 죄송합니다. 지면관계로 모든 참고문헌을 수록하지 못했음을 사과드립니다. 문헌요청시 기꺼이 응하겠습니다.