

## 재무비율을 이용한 부도예측에 대한 연구: 한국의 외부감사대상기업을 대상으로\*

박종원(교신저자)

서울시립대학교 경영대학 교수

(parkjw@uos.ac.kr)

안성만(주저자)

농협은행 여신정책부

(asm1230@nonghyup.com)

본 연구는 한국시장의 외부감사대상기업을 대상으로 재무비율을 이용한 부도예측모형을 개발하고 모형의 정확도와 변별력을 검증하였다. 상장기업들을 대상으로 이루어진 선행연구와 달리 외부감사대상기업 전체를 대상으로 재무비율을 이용한 부도예측 모형을 구축하고 그 적용가능성을 살펴보았다는 점에서, 기업의 부실을 설명해 줄 수 있는 변수를 파악하기 위해 구성가능한 117개의 대량의 재무비율을 사용하였다는 점에서 기존연구와 차별화된 특징을 갖는다. 모형의 변별력을 파악하기 위해 AUROC, AR, K-S 통계량 등을 사용하였다. 결산연도기준으로 2003년~2006년의 외부감사대상기업들의 데이터를 이용하여 다변량 로짓분석을 이용한 결과 금융비용대부채비율, 자기자본비율, 차입금의존도, 현금성자산비율, 총부채회전율, 영업이익대총자산비율, 총자본투자비율, 기업규모, 건설업 더미, 제조업더미 등 17개의 변수가 부도예측모형을 구성하는 변수로 선택되었다. 추정된 모형의 분류 정확도(concordant ratio)는 82%로 나타났으며, 부도예측에 대한 모형의 변별력 지표인 AUROC값은 85%, K-S 통계량은 55%로 매우 양호하게 나타났다. 이러한 결과는 데이터에 존재하는 극단치와 결측치에 영향을 받지 않는 결과이며, 매 연도별 분석에서도 큰 차이를 보이지 않아 시계열적으로도 강건한 결과이다. 이는 본 연구에서 구축한 모형이 한국시장의 외부감사대상기업의 부도예측에 유용하게 사용될 수 있음을 의미한다.

주제어: 외부감사대상기업, 부도예측, 재무비율, 다변량 로짓분석, AUROC, K-S통계량

### 1. 서론

한국 경제에 큰 타격을 준 1997년의 외환위기 이후 기업의 부도위험을 정확하게 판단하기 위한 노력이 지속적으로 이루어져왔으며, 2008년 미국시장의 서브프라임모기지 사태로 촉발된 세계금융위기 이후 그 중요성은 더욱 강조되고 있다. 기업의 부도위험을 정확하게 평가할 수 있다면 막대한 사회·경제적 손실을 야기하는 기업의 부도를 사전에 예측할 수 있고, 경영정책의 변화와 재무구조개선을 통해 기업

경영을 정상화시키는 동기를 제공할 수 있다. 또한 자발적인 청산의 경우에도 불필요한 자원의 낭비를 줄이고 자원의 효율적 배분을 가능케 함으로써 관련 비용을 최소화 시킬 수 있다는 점에서 부도예측을 위한 모형의 개발과 평가는 중요한 의미를 갖는다. 재무제표로부터 손쉽게 얻을 수 있는 재무비율은 기업부실을 예측하는 모형의 투입변수로 곧잘 활용된다. 특히 외부감사를 통해 신뢰성이 검증된 재무제표를 얻을 수 있는 외부감사대상기업(이하, 외감기업)의 경우 재무비율을 이용한 부도예측모형은 모형 구축의 용이성과 비용 측면에서 또 재무제표정보가

최초투고일: 2014. 2. 3      게재확정일: 2014. 3. 24

\* 이 논문은 2013년도 서울시립대학교 교내연구비 지원을 받아 작성되었으며, 두 번째 저자의 서울시립대학교 석사학위논문の内容 일부를 수정, 보완하여 발전시킨 것입니다.

갖는 정보의 예측력과 관련하여 실무적으로나 학문적으로 중요한 연구대상이다.

본 연구에서는 외부감사를 받는 기업을 표본으로 하여 다변량 로짓분석(multi-variate logistics regression)을 이용하여 기업부실을 설명할 수 있는 재무비율을 이용한 부도예측모형을 개발·평가하고, 외부감사대상기업의 부도위험 관리에 도움을 줄 수 있는 시사점을 도출한다. 상장기업을 대상으로 이루어진 국내 선행연구와 달리 본 연구는 우리나라의 외감기업 전체를 표본으로 분석하였다는 점에서 또 금융결제원의 당좌거래 정지 정보에 기초하여 포괄적인 부도기업표본을 구축하고 체계적인 데이터 정제과정을 거쳐 기업부실을 설명하는 재무비율을 분석하고 부도예측모형을 구축, 평가하였다는 데에서 큰 차별성을 갖는다.

본 연구에 사용된 표본은 2008년 세계 금융위기의 영향을 받기 전의 외감기업이다. 외부감사의견과 기업개요정보 및 재무제표는 한국신용평가정보(KIS)에서 제공하는 2001년부터 2006년까지의 데이터를 이용하였고 이를 바탕으로 생성한 재무비율을 기업부실에 대한 설명변수로 사용하였다. 부도정보는 2005년 1월 1일부터 2007년 12월 31일까지의 기간에 금융결제원에서 확인되는 당좌거래 정지(당좌수표정지 또는 약속어음부도) 외감기업을 이용하였다. 정상기업과 부도기업을 잘 구별하는 변수를 가려내기 위한 방법으로 T-검증, 단일변량 로짓분석, AR(accuracy ratio)을 이용하였으며, 다변량 분석 방법으로는 다변량 로짓분석을 이용하였다. 모형의 변별력을 검증하기 위한 방법으로는 AUROC(area under ROC), AR, Kolmogorov-Smirnov(K-S) 검증을 사용하였다.

본 연구의 주요결과를 요약하면 다음과 같다. 다변량 로짓분석의 결과 금융비용대부채비율, 유동부

채비율, 자기자본비율, 총자본투자효율 등 17개의 재무비율이 기업부실을 설명하는 중요 비율로 선택되었으며, 부도예측모형의 예측력을 분석한 결과 약 88%의 높은 AUROC값(76.2%의 AR값)을 보여준다. 등급 계량화를 위해 변수의 극단치와 결측치를 조정한 후의 AUROC값은 85%이다. K-S 통계량도 55.04%로 높게 나타나 구성된 부도예측모형의 변별력이 우수함을 보여준다.

본 논문은 구성은 다음과 같다. 제 2 장에서는 선행연구를 간략하게 살펴본다, 제 3 장에서는 본 연구에서 사용할 부도의 개념과 부도를 인식하는 기간을 정의하였으며, 분석 모집단을 선정하였다. 또한 데이터(재무제표와 부도정보)를 정제하는 과정을 기술하였으며, 모형에 사용할 변수를 정의하고 구체적인 연구방법론에 대하여 설명하였다. 제 4 장에서는 제 3 장에서 구성한 모집단을 이용해 단일변량 분석과 다변량 로짓분석을 이용하여 부도예측 모형을 구성하였으며, 모형에 대한 검증을 실시하였다. 마지막 제 5 장에서는 연구의 결과를 요약하고 연구의 시사점을 정리하였다.

## II. 선행연구

재무비율을 이용한 부도예측에 대한 연구는 Beaver(1966)의 단일변량 분석과 Altman(1968)의 다변량 판별분석(multi-variate discriminant analysis) 이후 다음의 두 방향으로 진행되고 있다. 하나는 기업의 상태에 대해 어떤 변수 또는 지표가 보다 많은 정보를 포함하고 있는가 하는, 정보원천에 관한 탐색이다. 주요 정보원천에는 재무정보와 같은 기업내부 지표, 기업경영 전반에 큰 영향을 미칠 수 있는

거시경제 지표, 주가 관련 지표들과 회계감사 결과와 같은 공식적인 외부평가 지표 등이 있다. 다른 하나는 모형 구축에 적용 가능한 다양한 통계적 기법에 대한 탐색이다. 정량적인 부도예측 모형은 Beaver(1966)의 이원분류검정에서 출발하여 Altman(1968)의 다변량 판별분석, 그리고 Ohlson(1980)의 로짓 모형, Zmijewski(1984)의 프로빗(probit)모형 등과 같은 이산확률모형으로 발전하여 왔고 최근에는 신경망(neural networks)모형, 생존분석모형 등 다양한 기법을 이용한 연구가 진행되고 있다.

단일변량 분석은 기업부실을 가장 잘 예측할 수 있는 하나의 변수를 이용하여 부도를 예측하는 모형으로 설명변수 상호간의 구조적 관계는 고려하지 않는다. 대표적 기법으로 프로파일분석과 이원분류법을 들 수 있다. 단일변량 분석은 적용의 간편성으로 인해 쉽게 이용될 수 있으나 여러 한계를 가지므로 다양한 비율을 복합적으로 이용하는 다변량 분석이 보다 널리 이용되고 있다.

다변량 분석은 독립변수 상호간에는 구조적 관계가 존재하며, 이들을 적절하게 조합할 경우 기업부실에 대한 예측력을 높일 뿐 아니라 개별변수가 모형에서 갖는 의미를 찾을 수 있다는 데에 그 논리적 기반을 두고 있다. 여러 변수를 동시에 고려하는 통계적 기법으로는 판별분석, 주성분분석, 그리고 다중회귀분석 등이 사용되는데 부도예측 모형의 경우 지금까지 판별분석과 로짓분석이 많이 이용되어 왔다.

판별분석은 계량적으로 측정된 독립변수들을 이용

하여 명목척도로 측정된 종속변수를 분류하는 방법으로 분류될 집단 간의 차이를 의미 있게 설명해 줄 수 있는 독립변수를 찾아내어 이들의 선형조합으로 다음과 같은 판별식을 만들어 내고, 이 식에 분류하고자 하는 각 대상의 독립변수를 대입하여 각 대상이 속하게 되는 집단을 찾아낸다.

$$Z_i = a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n$$

$Z_i$ : 기업  $i$ 의 판별점수(절사점)

$a_1, a_2, a_3, \dots, a_n$ : 판별계수

$X_1, X_2, X_3, \dots, X_n$ : 독립변수(재무비율)

판별분석에서 각 집단은 다변량 정규모집단에서 추출된 표본이며 모집단의 공분산행렬은 동일하다고 가정한다. 위의 판별함수에서 종속변수  $Z_i$ 는 판별점수를 나타내며, 독립변수  $X_1, X_2, X_3, \dots, X_n$ 는 판별변수이다. 추정된 판별함수 식에 각 기업이 갖는 판별변수의 값들을 대입하면 해당 기업에 대한 판별점수를 구할 수 있고, 이 판별점수에 의하여 집단을 구분하게 된다. 판별변수의 계수인  $a_i$ 는 집단 사이의 분산에 대한 비율이 최대화되도록 하는 상수이다. 다변량 판별분석은 부도예측은 물론 신용평가, 주식평가, 채권의 등급평가 등으로 그 적용범위가 점점 확대되고 있다.<sup>1)</sup>

최근 들어 로짓 회귀분석을 이용한 부도예측 연구가 많이 이루어지고 있다. 로짓 회귀분석은 설명하고자 하는 변수가 양분된 범주로 이루어져 있을 경

1) 다변량 판별분석을 이용한 부도예측의 대표적인 연구로 Altman(1968, 1983), Altman, Haldeman, and Narayanan(1977), 국내기업의 부도예측에 대한 연구로 전춘옥(1984), 강종만, 홍성의(1999), Altman(1996)을 참조할 수 있다. 그러나 판별분석은 다음과 같은 한계점을 지니고 있다. 첫째, 판별모형은 이미 발표된 과거의 재무자료를 사용하여 판별함수를 추정하고 이를 기준으로 미래 상황을 예측한다. 따라서 기업환경이 전체적으로 변화하거나 또는 경영패턴이 변하면 모형의 예측력이 떨어질 수 있다. 둘째, 어떤 재무변수가 판별함수를 추정하는데 최적인가에 대한 객관적 선정 기준과 이론적 근거가 명확하지 못하다. 셋째, 판별모형은 기업가치에 영향을 미치는 비회계적이거나 질적인 자료를 고려하지 못한다. 또한 표본기업들 사이에 회계처리 방법이 상이할 경우에는 판별함수의 예측력이 감소한다. 넷째, 판별분석에 의한 기업부실 판별모형은 판별집단의 예측치에 대한 분산-공분산행렬이 동일하고, 예측치의 분포가 정규분포를 갖는다는 가정을 충족해야만 하는 문제를 갖는다.

우에 주로 사용하는 방법으로, 종속변수가 비선형의 로짓 형태를 취하며 단지 두 개의 값만을 가지는 경우 종속변수와 설명변수 사이의 관계를 분석하는 통계기법이다. 판별분석과 로짓분석의 차이점을 살펴보면, 첫째, 판별분석은 독립변수들이 정규분포를 하며 집단 간에 공분산행렬이 동일하다고 가정하나, 로짓분석에서는 이러한 가정에 위배되어도 선택확률이 로짓 함수를 취한다는 가정 하에서 사용할 수 있다. 둘째, 판별분석의 판별점수가 단순한 순위만을 나타내는 반면 로짓모형의 선택확률  $P$ 는 0과 1 사이에 존재하여 절사점으로서 부도 또는 정상의 판별은 물론 기업의 부실가능성을 나타내는 확률을 제시함으로써 정보의 질을 높일 수 있다. 셋째, 판별모형의 경우 계수의 유의성 검증이 불가능한 반면, 로짓모형에 있어서 독립변수들의 각 계수는 유의성 검증이 가능하여 각 독립변수와 부도확률과의 상관관계를 판단할 수 있게 한다.

위에서 설명한 것과 같이 로짓분석은 선형회귀분석과 유사하게 적용할 수 있고, 비선형적인 효과를 통합하고 있으며 기업부실에 대한 전반적인 진단을 내릴 수 있다는 유용성으로 인해 부도예측 관련연구에 유용하게 사용되고 있다.<sup>2)</sup> 다음에서는 로짓분석을 이용한 부도예측에 대한 주요 선행연구를 간단하게 살펴본다.

Ohlson(1980)은 로짓분석을 이용하여 1970년

부터 1976년까지의 105개 부실기업과 2,058개의 건전기업을 표본으로 부실 1년 전과 2년 전 자료로부터 9개의 재무비율을 선정, 분석한 결과 총자산 대 물가지수 비율, 총부채 대 총자산 비율, 운전자본 대 총자산 비율, 순이익 대 총자산 비율, 영업활동에서 조달된 운전자본 대 총부채 비율 등 다섯 가지 변수가 중요한 부도예측 변수임을 확인하였다. Zavgren(1985) 역시 로짓분석을 이용하여 1972년부터 1978년까지의 45개 부실기업과 45개의 건전기업을 표본으로 선정하고, 부실 전 5년간의 재무자료로 5년 후까지 예측할 수 있는 5개의 부도예측모형을 구성하였다. 모형 구축에 포함된 재무비율은 매출채권/재고자산, 재고자산/매출액, 현금/총자산, 당좌자산/유동부채, 순이익/총자산, 총부채/총자산, 매출액/유형고정자산 등 7개이며, 이 중 총부채/총자산은 5개년도 전체 모형에서, 재고자산/매출액은 도산 3년 전부터 5년 전까지의 모형에서, 당좌자산/유동부채는 도산 1년 전부터 3년 전까지의 모형에서 유의한 예측력을 가짐을 보여주었다.

송인만(1987)은 1978년에서 1986년 사이에 부실화된 35개 부실기업과 이와 동일 업종으로서 자산규모가 비슷한 35개 정상기업을 표본으로 추출한 후 매출액순이익률, 총자본회전율, 자기자본비율 등 3가지 재무비율을 독립변수로 선정하여 로짓분석을 통해 예측모형을 도출하였다. 그는 총자본회전율과

2) 프로빗(probit)분석 역시 로짓분석과 같이 설명하고자 하는 변수가 양분된 범주로 이루어져 있을 경우에 주로 사용하는 분석방법이다. 구체적으로는 독립변수와 종속변수의 관계에 특정한 누적분포함수를 도입한 기법으로서 로짓(logit)은 표준로짓분포 함수를 사용한 것인 반면 프로빗(probit)은 단위정규분포 함수를 이용한다. 즉, 프로빗모형은 개별 독립변수들이 독립적으로 특정한 분포를 따르고 이들 분포들이 결합된 결합확률분포에 의해 부도확률이 결정됨을 가정하고 있다. Zmijewski(1984)는 Ohlson(1980)의 로짓모형을 개선한 프로빗모형을 제시하였다. 신경망분석은 인간의 두뇌가 문제를 해결하는 구조를 이용한 분석으로 기업의 부실원인의 성질이 비선형임을 감안할 때 적절한 것이다. 이모형을 이용한 부도 예측은 기업의 각종 회계자료 및 재무지표를 신경망모형에 입력, 훈련시켜 부도가능성을 예측하는 기법이다. 1980년대 후반부터 신경망 기법들이 부실예측모형 연구에 많이 활용되고 있다. 이진창(1993)은 다변량 판별분석, 귀납적 학습방법, 인공신경망 등 3가지 모형의 예측력과 성과를 비교하였다. 인공신경망은 다시 다변량 판별분석에 의해 선택된 변수를 이용한 인공신경망 I 과 귀납적 학습방법에 의해 선택된 변수를 이용한 인공신경망 II로 구분하여 모두 4가지 모형으로 비교 분석한 결과, 인공신경망 II > 인공신경망 I > 귀납적 학습방법 > 판별분석의 순으로 높은 예측력을 보인다는 결과를 제시하고 있다.

자기자본비율이 기업부실을 예측하는 가장 유의한 변수임을 보여준다. 이계원(1993)은 기업부실의 발생원인과 결부시켜 예측모형 구축을 위한 변수를 선정하고 로짓분석을 통해 부도 전 3개 연도별로 모형을 추정하였다. 부실기업 표본은 1985년 1월부터 1992년 7월까지 부실화 된 41개 상장기업이며, 여기에 1:3의 비율로 대응하여 123개의 건전기업을 추출하여 분석을 실시하였다. 연구 결과 자기자본비율, 현금흐름비율, 비기대재고자산비율은 부실 3년 전부터 부실예측에 통계적 유의성이 있으며 기업규모는 부실 1년 전에 통계적으로 유의한 설명력을 가진다는 결과를 제시하였다. 장휘용(1998)은 1991년부터 1994년까지의 기간에 부도가 발생한 34개 기업과 건전기업 103개를 표본기업으로 선정하고, 총자산경상이익률, 주당순이익변화율, 재고자산회전율, 매출채권회전율, 총자산회전율의 변화율, 매출액대비금융비용비율, 주당장부가치 등 7개 변수로 부도예측모형을 구축하였다. 박정윤(2000)은 1991년부터 1996년까지 부실이 발생한 상장기업 41개사와 118개의 건전기업을 표본으로 선정하고, 자기자본순이익률, 총자본부채비율 및 배당률, 현금흐름변수를 포함하여 로짓분석을 실행한 결과, 부실 4~5년 전에는 자기자본순이익률과 총자본부채비율이 기업부실에 유의한 영향을 주나, 부실 전 3년간은 현금흐름과 배당률의 크기가 부실에 영향을 준다는 결과를 보여준다.

다음의 <표 1>은 부도예측과 관련한 대표적인 선행연구 내용을 간략하게 요약한 것이다.

### III. 연구방법론

#### 3.1 부도의 정의와 인식

##### 3.1.1 부도의 정의

부도를 명확하게 정의하고 부도시점을 정확하게 인식하는 것은 부도예측에 대한 연구에서 매우 중요하다. 현재 국내에서 부도기업을 파악하기 위해 사용하는 부도에 대한 정의는 금융감독원의 정의, 신BIS협약의 정의, 은행연합회의 불량정보기준 등이 대표적이다.<sup>3)</sup> 이 중 신BIS협약의 부도 정의는 기업의 잠재적 부실가능성을 포함하여 가장 포괄적으로 부도를 정의한다. 은행연합회의 불량정보기준은 국내 대부분의 금융기관에서 사용하는 기준으로 신BIS협약에서 포괄적으로 정의하고 있는 부도에 해당하는 사유를 구체적으로 제시하고 있다. 은행연합회의 불량정보 기준을 이용할 경우 은행에서 사용하는 부도 차주를 대상으로 분석을 하기 때문에 보다 실제적인 연구를 수행할 수 있다는 장점이 있으나, 데이터를 입수하기 어렵다는 문제점이 있다. 따라서 본 논문에서는 금융결제원에서 제공하는 당좌수표정지 또는 약속어음부도로 인해 당좌거래가 정지된 기업을 부도기업(또는 불량기업)으로 정의하고 분석에 사용한다.<sup>4)</sup>

##### 3.1.2 부도의 인식기준과 재무제표 정보

부도를 어느 시점에서 인식하느냐에 따라 실제 부

3) 금융감독원 신BIS실, 2007, 6, “위험가중자산에 대한 자기자본비율 산출기준”, 은행연합회 신용정보관리규약 참조.

4) 금융결제원 당좌거래정보, <http://www.kftc.or.kr/kftc/info/dishonor/period.jsp>. 이 기준은 한국은행에서 부도기업을 집계하는 데에 사용된다.

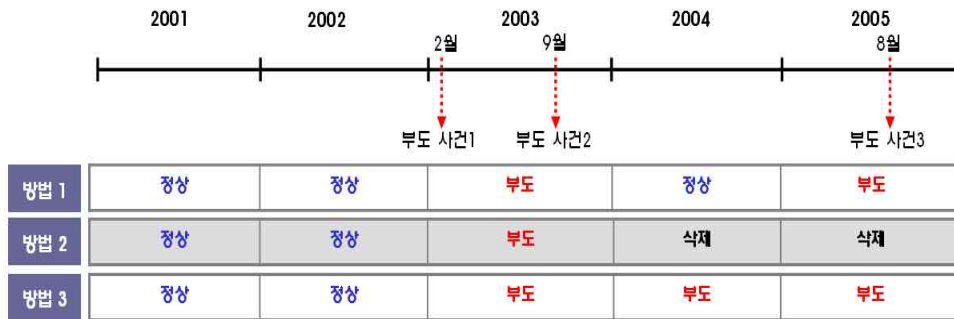
〈표 1〉 주요 선행연구의 요약

연구자	연도	표본기업	연구방법	연구결과
Beaver	1966	상장 기업 부도 : 79개 정상 : 79개 부도율 : 50%	단일별량 분석	이원분류검정을 통한 분석결과 현금흐름/총부채비율의 예측력은 부도 1년, 5년 전에 각각 87%, 78%로 가장 좋은 예측지표로 밝혀졌으며, 그 다음으로는 총자산순이익률의 예측력이 높게 나타남.
Ohlson	1980	상장 제조업 부도 : 105개 정상 : 2,058개 부도율 : 4.85%	로짓분석	9개의 재무변수를 이용하여 예측모형을 구축, 분석한 결과 기업구조, 재무구조를 나타내는 변수, 수익성 및 유동성을 나타내는 변수가 부실확률모형에 유용한 것으로 확인됨.
전춘옥	1984	상장 제조업 부도 : 29개 정상 : 29개 부도율 : 50%	다변량 판별분석	31개 재무비율과 3개의 실수 중 통계적 유의성이 있는 지표로 매년 4개의 주성분을 판별함수로 도출한 결과 부도 1년, 2년, 3년 전에 각각 94.83%, 77.59%, 75.86%의 예측력을 보임.
이계원	1993	상장 기업 부도 : 34개 정상 : 123개 부도율 : 22%	로짓분석	ROE변수만을 사용한 기본모형과 여기에 변수를 추가한 확장모형으로 연구한 결과 기본모형의 예측력은 부도 1년 전의 경우 예측정확성은 80%였으나 확장모형은 82%로 다소 높게 나타남.
Altman	1996	상장제조/도소매 부도 : 34개 정상 : 61개 부도율 : 36%	판별분석	K1-Score 모형의 예측력은 부도 전 5년간 71.1%, 88.2%, 69.7%, 50.0%, 68.8%로 나타났으며, K2-Score 모형의 경우에는 96.6%, 85.2%, 71.4%, 40.0%, 75.0%로 나타남.
장휘용	1998	상장 기업 부도 : 34개 정상 : 103개 부도율 : 25%	로짓분석	설명변수 7개로 모형을 구축하였는데, 자체예측력은 부실표본에서 85.3%, 정상표본에서 95.1%였으며 검정용표본에서는 부실과 정상표본에서 각각 76.5%와 94.2%의 예측력을 보임.
박정운	2000	상장 기업 부도 : 41개 정상 : 118개 부도율 : 26%	로짓분석	기업의 주요 재무정책변수 중 부실 4~5년 전에는 자기자본순이익률과 총자본부채비율이 부실에 영향을 주나, 부실 전 3년간은 현금흐름과 배당률의 크기가 부실에 영향을 주는 것으로 나타남.

도율이 과대 또는 과소 산출될 수 있으므로 이에 대한 엄밀한 기준이 필요하다. 본 논문에서는 〈그림 1〉에서의 방법2를 적용하여 최초 부도시점을 부도로 인식하고 부도 이후에 대하여는 회복여부에 대하여 고려하지 않고 분석 대상에서 제외한다.

〈그림 1〉에서 방법 1은 회복을 인정하고 부도 사건이 발생하지 않은 연도에는 정상 차주로 인식하는

방법이다. 방법 2는 최초 부도시점을 부도로 인식하며, 그 이후에는 회복된다 하더라도 부도 차주로 간주하고 분석대상에서 삭제하는 방법이다. 이는 부도 차주의 부도 사유가 해제되었다 할지라도 실질적인 회복 여부에 대하여 차주별로 확인이 어려워, 회복된 차주를 정상적인 차주로 보기 어려운 동시에 부도 차주로 판단하기에도 무리가 있기 때문이다. 방



〈그림 1〉 부도 인식 기준

법 3은 최초 부도시점을 부도로 인식하고 그 이후에는 매년 불량집단으로 인식해서 모집단에 포함시키는 방법인데, 이는 부도 업체수가 상당히 부족할 때 사용할 수 있다. 특정 기업에 대해 동일 연도에 부도 사건이 중복 발생하는 경우는 그해 최초로 부도사건이 발생한 날짜를 부도 발생일로 설정하였다.

일반적으로 재무비율을 이용한 부도예측 및 평가모형을 개발할 경우 부도 관측기간을 결산(12월말 결산법인 가정) 재무제표가 공시되는 시점을 고려하여 당해년도 4월부터 이듬해 3월까지 12개월을 부도 관측기간으로 정의한다. 왜냐하면 결산 재무제표가 공시되는 시점을 고려하지 않을 경우 부도의 원인이 되는 사건이 발생한 기간의 결산자료를 이용해 부도를 예측하는 오류를 범할 수 있기 때문이다. 예를 들어, 2002년 12월말이 결산일인 외감기업의 경우 2003년 3월말 이후에 해당기업의 재무정보를 확인할 수 있다. 2002년의 재무정보는 2003년의 부실여부를 사전에 예측하기 위하여 이용되는데, 그 정보는 2003년 4월 이후에 알 수 있다. 따라서 결산시점과 예측시점의 시차를 고려하지 않으면 재무

상태 악화가 이미 반영된 2003년의 재무정보를 이용해 2003년 중의 부도여부를 예측하는 오류를 범하게 된다.<sup>5)</sup>

하지만 이론적으로는 2002년의 재무정보를 연말에 알 수 있다는 가정 하에서 2003년 1월 1일부터 12월 31일까지의 불량여부를 판단하는 재무정보는 직전년도 재무제표에서 얻는 정보이다. 따라서 본 논문에서는 부도인식 기간을 재무제표기간(1월 1일~12월 31일)과 동일하게 정의하였다. 예를 들어 2003년 2월 중에 부도가 발생한 경우 2002년 재무정보가 평가 대상이 되는 것이다.<sup>6)</sup> 즉, 2003년에 부도 사건이 발생한 경우 전년도(2002년) 재무제표를 부도재무제표로 파악하여 분석에 사용하였다. 하지만 부도 사건 직전년도 재무제표가 존재하지 않는 경우 부도데이터 확보에 어려움을 생길 수 있기 때문에 전전년도(2001년) 재무제표를 사용하였다. 이러한 이유는 실제로 부도가 발생할 경우 직전연도의 재무제표를 이용할 수 없는 경우가 많고, 이 경우 부도의 신호는 전전년도 재무제표에서도 나타난다고 할 수 있기 때문이다.

5) 2002년 재무제표가 공시되기 전인 2003년 2월 20일에 관리대상으로 지정되는 경우, 현실적으로 2001년의 재무정보가 불량해서 관리대상으로 지정된다.

6) 실제 운영시라면 2002년의 재무제표는 공시전이기 때문에 2001년의 재무제표를 이용할 수밖에 없다.

〈표 2〉에 상황에 따른 다양한 부도인식의 예를 나타내었다. 〈표 2〉에서 기업 A의 경우 부도사건이 2006년에 발생하였고 부도사건 발생시점에서 2005년도 결산재무제표가 없는 경우이다. 따라서 2004년도 결산재무제표를 부도표본으로 이용한다. 기업 B와 C는 부도 시점에서 직전년도 재무제표를 확인할 수 있는 경우이다.

### 3.2 표본 선정과 데이터의 정제

#### 3.2.1 표본 구성을 위한 기초자료

연구에 사용될 표본 선정을 위하여 금융기관과 공공기관을 제외한 기업 중 공인회계사의 외부감사의견이<sup>7)</sup> 있는 12월 말 결산법인인 17,913개 기업의 2001년 1월 1일부터 2006년 12월 31일까지 결산 재무제표를 이용하였다. 부도기업 표본구성을 위한 부도정보는 2005년 1월 1일부터 2007년 12월 31일까지의 금융결제원 당좌거래정보를 이용하였으며,

당좌수표정지 또는 약속어음부도로 인해 당좌거래가 정지된 기업 차주들을 대상으로 선별하였다. 전체 17,913개 기업 가운데 322개 기업에서 부도가 발생하였다.

#### 3.2.2 데이터의 정제와 분석대상 표본

데이터 정제(精製, cleansing)란 불완전하고 오류가 있는 데이터를 보완/수정하는 일련의 과정을 말한다. 원시 데이터(raw data)를 이용해서 바로 분석하지 않고 정제과정을 거치는 이유는 다양한 오류(프로그램 오류, 수기 입력 오류, 매핑(mapping) 오류 등)으로 인한 왜곡된 데이터로 분석을 실시할 경우 사실과는 다른 결과를 도출해 낼 수 있기 때문이다. 본 연구에는 〈그림 2〉에서 보는 것과 같은 6단계의 정제과정을 거쳐 분석대상 표본을 구성하였다(구체적인 정제과정은 부록에 정리하였다).

6단계의 정제과정을 거쳐 최종적으로 구성된 표본은 제조/건설/도소매/서비스업에 포함되는<sup>8)</sup> 외감기

〈표 2〉 부도 인식의 예

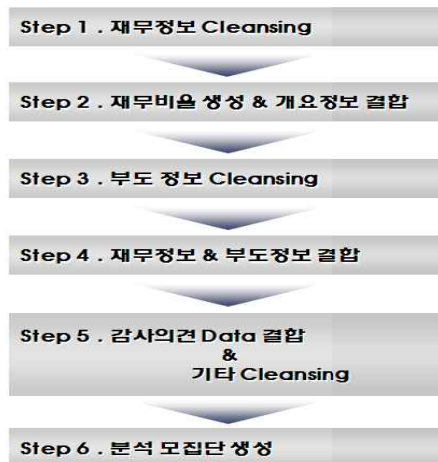
업체명	결산연도 재무제표	최초 부도사건 발생 년도	부도여부
A	2002	2006	정상
A	2003	2006	정상
A	2004	2006	부도
B	2003	2005	정상
B	2004	2005	부도
C	2004	2007	정상
C	2005	2007	정상
C	2006	2007	부도

7) 본 연구에서는 외감기업을 자산규모가 70억 이상이 아니더라도 실질적인 외부감사의견이 있는 기업으로 정의하였다.

8) 기업은 규모(size)에 따라 부도위험이 구분되기도 하지만 해당기업이 속한 산업(업종)에 따라서도 부도위험은 현저하게 다를 수 있다. 따라서 표본기업들을 산업별로 그룹화 하여 적절한 규모로 분류하는 것이 연구에 도움이 될 수 있다. 본 연구에서는 부도기업의 수가 많지 않기 때문에 산업구분은 제8차 한국표준산업분류(KSIC) 기준에 따라 제조업/건설업/도소매업/서비스업으로 구분하고, 이후 분석은 제조업과 건설업/도소매업/서비스업으로 구분하여 시행하였다.



업 중 2003년 ~ 2006년의 결산연도 재무제표를 이용할 수 있는 12월 말 결산법인들이다. 표본을 구성하는 총 기업-연도 수는 정상기업과 부도기업을 포함해 모두 52,052개로 대부분의 외감기업을 포함하고 있으며, 이중 부도기업은 322개로 한국은행에서 부도로 정의하는 2005년 1월 1일부터 2007년 12월 31일까지의 부도기업을 대부분 포함하고 있다.



〈그림 2〉 데이터의 정제와 표본 구성 과정

〈표 3〉은 산업별 표본의 구성내역을 나타낸다. 〈표 3〉에서 보듯이 연도별 표본기업의 수는 11,466개에서 14,464개의 분포를 보이며, 그 중 50% 정도는 제조업에 속한다. 전체 표본을 결산연도와 산업별로 구분해 보았을 때 결산연도 재무제표 기준으로 2004년과 2005년에(실제 연도는 2005년과 2006년이 될 것임) 부도가 많이 발생하였고 제조업이 타 업종(건설/도소매/서비스업) 보다는 부도율이 높다는 것을 알 수 있다. 다음으로 건설업의 부도율이 높고 서비스업, 도소매업 순이다.<sup>9)</sup> 매 연도별 실제 부

도율은 0.3%에서 0.86%의 분포를 보이며 평균적으로는 0.62%의 값을 갖는다.

### 3.3 변수의 선정

본 연구에서 사용한 주요 변수는 재무제표에서 구한 재무비율 변수이다. 총 117개의 재무비율을 사용하여 분석을 실시하였는데, 재무비율을 〈표 4〉와 같이 9개의 범주로 나누었다. 재무비율에 대한 정의와 계산 방법은 한국은행에서 발표하는 기업경영분석을 따랐다. 전체 재무비율에 대한 상세내역 보고는 생략한다.

연구에 사용될 독립변수는 다음과 같은 과정을 통해 선정한다. 먼저 각각의 재무비율 값이 부도기업과 정상기업 간에 유의적인 차이가 있는지를 117개의 후보 재무비율에 대한 T-검증 / 단일변량 로짓 / Accuracy Ratio(AR) 등의 방법으로 변별력을 확인하여 유의한 재무지표를 선택한다. 사용된 유의성 기준은 다음과 같다.

- T-검증 :  $\text{Prob}(t) < 0.05$
- 단일변량 로짓 :  $\text{Prob}(\text{Chisq}) < 0.05$
- $\text{AR} \geq 30\%$

이후 단계적 선택법(stepwise selection method)을 통하여 기업의 부도 예측지표로 유효하다고 판단되는 비율들을 선정하여 다변량 로짓모형의 설명변수로 사용한다.

9) 2003년과 2004년은 도소매업이 서비스업보다 부도율이 높게 나타났다.

〈표 3〉 산업별 표본의 구성내역

년도는 결산연도 재무제표 기준임.

년도	산업구분	정상	부도	전체	구성비	부도율
2003	제조업	5,962	20	5,982	11%	0.33%
	건설업	1,110	8	1,118	2%	0.72%
	도소매업	1,119	2	1,121	2%	0.18%
	서비스업	3,241	4	3,245	6%	0.12%
	소계	<b>11,432</b>	<b>34</b>	<b>11,466</b>	<b>22%</b>	<b>0.30%</b>
2004	제조업	6,237	78	6,315	12%	1.24%
	건설업	1,291	17	1,308	3%	1.30%
	도소매업	1,197	4	1,201	2%	0.33%
	서비스업	3,771	10	3,781	7%	0.26%
	소계	<b>12,496</b>	<b>109</b>	<b>12,605</b>	<b>24%</b>	<b>0.86%</b>
2005	제조업	6,648	60	6,708	13%	0.89%
	건설업	1,371	23	1,394	3%	1.65%
	도소매업	1,289	2	1,291	2%	0.15%
	서비스업	4,109	15	4,124	8%	0.36%
	소계	<b>13,417</b>	<b>100</b>	<b>13,517</b>	<b>26%</b>	<b>0.74%</b>
2006	제조업	7,006	40	7,046	14%	0.57%
	건설업	1,508	27	1,535	3%	1.76%
	도소매업	1,410	2	1,412	3%	0.14%
	서비스업	4,461	10	4,471	9%	0.22%
	소계	<b>14,385</b>	<b>79</b>	<b>14,464</b>	<b>28%</b>	<b>0.55%</b>
2003 ~ 2006	제조업	25,853	198	26,051	50%	0.76%
	건설업	5,280	75	5,355	10%	1.40%
	도소매업	5,015	10	5,025	10%	0.20%
	서비스업	15,582	39	15,621	30%	0.25%
	총합계	<b>51,730</b>	<b>322</b>	<b>52,052</b>	<b>100%</b>	<b>0.62%</b>

〈표 4〉 범주별 재무비율 건수

범주	재무비율 수
부채상환능력	16
비용구조	4
생산성	5
성장성	10
수익성	14
안정성	18
유동성	10
현금흐름	22
활동성	18
합계	117

### 3.4 다변량 로짓분석

본 연구에서는 어떤 요인(독립변수)들이 부도(종속변수)의 위험요소이며, 각 요인들이 기업의 부도에 어떤 영향을 끼치는가를 살펴보기 위해 다변량 로짓분석을 이용한다. 다음에 로짓 회귀분석에 대해 간략히 설명한다.

#### 3.4.1 로짓 함수와 로짓 회귀분석

종속변수가 0(정상)과 1(부도)의 값을 갖는 범주

형 변수인 경우에, 종속변수  $y$ 의 기댓값  $E(y)$ 가 설명변수  $x$ 가 증가함에 따라 S자 모양의 곡선을 그리며 서서히 1로 수렴하는 양상을 보이는 함수를 로짓 함수(logistic function)라 부르고,  $z = \beta_0 + \beta_1 x$ 라 놓을 때 종속변수의 기댓값은 식 (1)과 같이 표현되며 이 값은 어떤 사건(부도)이 발생할 확률로 해석할 수 있다.

$$E(y) = \frac{\exp(z)}{1 + \exp(z)} \quad \text{or} \quad \frac{1}{1 + \exp(-z)} \quad (1)$$

로짓 회귀분석(logistic regression)이란 단지 두 개의 값만을 가지는 종속변수와 독립변수들 간의 인과관계를 로짓 함수를 이용하여 추정하는 통계기법이다. 로짓 함수는  $\beta_0$ 와  $\beta_1$ 에 대하여 비선형함수이나 이를 선형으로 변환시킬 수 있다. 로짓 함수의 기대반응  $E(y)$ 는 확률을 의미하므로,  $E(y) = p_x$ 로 놓으면 다음과 같은 변환을 통하여 선형화된다.

$$\ln\left(\frac{p_x}{1-p_x}\right) = z = \beta_0 + \beta_1 x \quad (2)$$

식 (2)와 같은 변환을 로짓 변환(logistic transformation)이라고 부른다. 식 (2)에서 좌변의  $p_x/(1-p_x)$ 를 오즈(odds)<sup>10)</sup>라 한다. 독립변수의 수가 두 개 이상인 경우에도 로짓 회귀모형이 가능하다. 이 경우에는  $z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$ 로 나타내고, 로짓 함수는 식 (3)과 같이 추정된다.

$$\hat{p}_x = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k)} \quad (3)$$

따라서 적절한 기준치(cut-off point)  $c$ 를 정하여 판별규칙을  $\hat{p}_x \geq c$ 이면 부도로 분류하고  $\hat{p}_x < c$ 이면 정상으로 분류할 수 있다.

#### 3.4.2 결과 값 산출 방법과 분류의 정확도

로짓 모형의 결과 값으로 산출된  $\hat{p}$  값에 대해 두 집단의 표본으로 짝을 지어 Pairs를 생성하면, 총 (우량집단 개수  $\times$  불량집단 개수) 만큼의 Pairs가 생성된다. 이후 다음과 같은 과정을 거쳐 Concordant, Discordant, Tied의 값을 산출하게 된다. 첫 번째는 각 Pairs에서 우량  $\hat{p}$ 과 불량  $\hat{p}$  값을 비교하여 (우량  $\hat{p} >$  불량  $\hat{p}$ ) 경우 0을 부여한다.<sup>11)</sup> 두 번째로 각 Pairs에서 우량  $\hat{p}$ 과 불량  $\hat{p}$  값을 비교하여 (우량  $\hat{p} =$  불량  $\hat{p}$ ) 경우 0.5를 부여한다. 마지막으로 각 Pairs에서 우량  $\hat{p}$ 과 불량  $\hat{p}$  값을 비교하여 (우량  $\hat{p} <$  불량  $\hat{p}$ ) 경우 1을 부여한다. 이렇게 해서,

$$\text{Concordant}(\%) = \frac{\text{1인 pairs 개수}}{\text{총 pairs 개수}}$$

$$\text{Discordant}(\%) = \frac{\text{0인 pairs 개수}}{\text{총 pairs 개수}}$$

$$\text{Tied}(\%) = \frac{\text{0.5인 pairs 개수}}{\text{총 pairs 개수}}$$

의 값들을 산출하게 된다.

10) 오즈는 확률과 관련된 의미를 갖는다. 예를 들어 스포츠게임에서 한 팀이 다른 팀을 이길 오즈가 4라는 말은 이길 확률이 4배라는 의미이다. 이길 확률  $p = 4/(1+4) = 0.8$ 이 된다. 즉,  $0.8/(1-0.8) = 4$ 이다.

11)  $\hat{p}$ 는 부실 확률이다.

여기서 각 값들을 살펴보면, Concordant(%)는 총 검증표본 중에서 표본내의 2개 집단에 대하여 우량을 우량으로, 불량을 불량으로 정확히 분류한 표본수가 총표본수에서 차지하는 비율이고 Discordant(%)는 총 검증표본 중에서 표본내의 2개 집단에 대하여 우량을 불량으로, 불량을 우량으로 잘못 분류한 표본수가 총표본수에서 차지하는 비율을 뜻한다. 그리고 Tied(%)는 총 검증표본 중에서 표본내의 2개 집단에 대하여 분류 정확도가 50%인 (표본 수/총 검증표본 수)를 나타낸다. 따라서 Concordant 값이 높고 Discordant 값이 낮을수록 우량(건전)집단과 불량(부도)집단 사이의 분류의 정확도가 높다고 할 수 있다.

### 3.5 모형의 변별력 평가방법

부도예측모형의 변별력이란 실제 부도기업을 부도로, 정상기업을 정상으로 얼마나 정확히 예측하였는가를 의미한다. 특정 모형에 의한 평가치를 이용하는 경우, 이 모형에 의해 평가되는 신용등급이 낮을수록 부도기업으로 판정될 가능성이 높을 것이다. 그리고 부도기업으로 판정될 가능성이 높을수록 실제 부도기업일 가능성이 높으면 그 모형은 타당한 모형이라 할 수 있다. 따라서 본 연구에서는 모형에 의해 추정된 등급이 낮을수록 우선적으로 부도기업으로 판정된다는 가정 하에, 전체기업 가운데 부도 판정을 받은 기업의 비율별로 각 상태(부도 또는 정상)의 상대빈도가 어떻게 달라지는가를 살펴봄으로써 특정한 부도 판정기준을 가정하지 않고 모형의 변별력을 파악하는 방법을 사용한다.<sup>12)</sup>

부도여부에 대한 사전적 판정기준을 가정하지 않고 각 상태의 발생빈도를 파악하는 데에는 발생가능한 모든 판정기준 가운데 임의의 판정기준 하에서 발생하는 각 상태의 빈도를 나타내는 점들을 연결한 곡선을 이용한 분석이 이용된다. 이러한 분석은 X축의 값이 전체 기업 중 부도로 예측된 기업이 차지하는 비중을, Y축의 값이 실제 부도기업 중 부도로 예측된 기업이 차지하는 비중을 나타내는 곡선을 그려봄으로써 손쉽게 이루어질 수 있다. 이러한 곡선을 CAP(cumulative accuracy profile) 곡선이라 한다.<sup>13)</sup>

한편, 모집단의 부도율 특성에 따라 모형에 의해 부도로 판정된 기업 중 실제 부도기업이 차지하는 비중인 변별력은 달라질 수 있다. 예를 들어, 부도율이 10%인 모집단을 대상으로 10%의 기업이 부도날 것으로 판정하는 경우, 부도로 판정된 기업 중 실제 부도기업이 차지하는 비중의 최댓값은 100%이다. 하지만 부도율이 20%인 경우, 동일하게 10%의 기업이 부도날 것으로 판정한 경우에는 그 최댓값이 50%이다. 따라서 전체 기업 집단의 특성에 상관없이 모형의 변별력을 판단하기 위해서는 전체 기업의 부도빈도와 관계없이 실제 정상인 기업들 중 부도로 예측된 기업이 차지하는 비중을 구해 볼 필요가 있으며, 이러한 곡선을 ROC(receiver operating characteristics) 곡선이라고 한다. 이는 CAP곡선과 Y축은 동일하지만 X축에 실제 정상인 기업들 중 부도로 예측된 기업이 차지하는 비중을 나타낸다는 차이가 있다.

모형의 변별력을 파악하는 지표로 CAP 곡선에서 계산되는 값이 AR(accuracy ratio)이며, ROC곡

12) 기존의 여러 연구들은 실제 부도 여부에 따라 기업의 상태를 구분하고, 모형에 의한 부도 예측결과와 실제 부도 여부가 어느 정도 일치하는가로 변별력을 계산하는 방법을 이용한다.

13) CAP Curve는 Gini Curve, Power Curve, Lorenz Curve 라고도 불린다.

선에서 구해지는 지표가 AUROC(area under ROC)이다. 두 지표를 이용하여 모형의 변별력을 판단할 수 있지만, 모형이 통계적으로 유의한 변별력을 갖는지는 판단할 수 없다. 이러한 단점을 보완하기 위해 Kolmogorov-Smirnov(K-S) 통계량을 사용한다. 따라서 본 연구에서 사용한 모형의 변별력 평가 방법은 AUROC, AR, K-S 통계량이다. 이에 대해 보다 자세히 살펴본다.

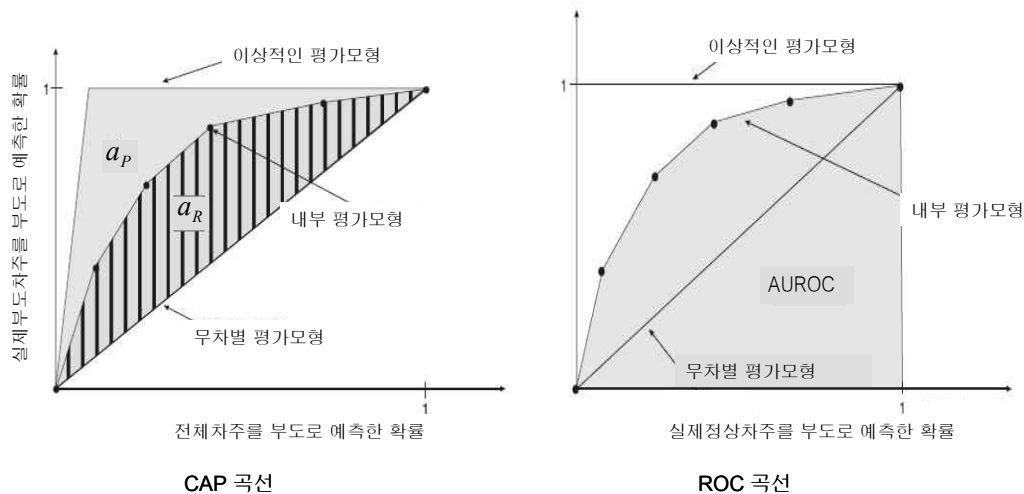
### 3.5.1 CAP, ROC, AR, AUROC<sup>14)</sup>

정상기업은 높은 등급을, 부도기업은 낮은 등급을 갖도록 모형이 개발되었고, 모든 기업이 모형의 등

급에 따라 낮은 등급부터 순서대로 나열되었다고 가정하고 CAP곡선과 ROC곡선을 이용하여 모형의 변별력을 어떻게 측정할 수 있는지 살펴보자.

CAP곡선과 낮은 등급에서 높은 등급까지 움직이면서 “전체기업 중 부도로 예측한 확률”과 “부도기업을 부도로 예측한 확률”을 비교하며, ROC곡선은 “정상기업을 부도로 예측한 확률”을 “부도기업을 부도로 예측한 확률”과 비교함으로써 변별력을 측정하는 것이다.<sup>15)16)</sup>

구체적 내용을 살펴보자. <그림 3>의 왼편에 나타난 CAP곡선에서 X축은 전체 기업 중 부도로 예측된 기업이 차지하는 비중이며 Y축은 실제 부도기업 중 부도로 예측된 기업의 비중이다. 그림에서 어떠



<그림 3> CAP 곡선과 ROC 곡선

14) 임종건, 2005, “신용평가시스템에 대한 적합성 검증 방법론 소개”, 금융감독원.

15) CAP곡선은 2002년 이후 Moody’s KMV에서 자체적으로 개발한 부도 예측모형인 RiskCalC<sup>TM</sup>의 변별력을 측정하는 도구로 사용된 이후 널리 알려졌으며, ROC 곡선은 의학 분야 등에서 새로 개발된 시약이나 치료방법의 성과 등을 비교하는 데 주로 사용되었던 방법이다.

16) CAP곡선은 “How much of an entire portfolio would a model have to exclude to avoid a specific percentage of defaulters?”라는 질문에 대한 답을 보여 주지만, ROC곡선은 “What percentage of non-defaulters would a model have to exclude a specific percentage of defaulters?”에 대한 답을 보여 준다. Roger M. Stein (2002), *Benchmarking Default Prediction Models*, Moody’s KMV 참조.

한 변별력도 존재하지 않는 무차별 모형(random model)은 원점과 (1,1)점을 지나는 기울기 45°의 직선으로 나타난다. 예를 들어 100개의 기업(정상기업: 90개, 부도기업: 10개)이 낮은 점수를 기준으로 정렬되어있을 때, 전체 기업 중 10%에 해당하는 점수 이하를 받은 기업을 모두 부도기업으로 판단한다고 하자. 변별력이 완벽하다면, 하위 10% 점수를 받아 부도기업으로 예측된 10개의 기업이 모두 실제 부도기업일 것이고, 아무런 변별력이 없다면 10개의 부도기업이 랜덤하게 선택될 것이므로 평균적으로 부도를 10%에 해당하는 1개의 기업이 실제 부도기업일 것이다. 따라서 특정 모형의 CAP곡선은 <그림 3>의 왼편 그림에서와 같이 변별력이 완벽한 이상적인 모형과 아무런 변별력이 없는 무차별 모형 사이에 나타나게 되며, 그 선이 이상적인 모형에 가까울수록 변별력이 높음을 나타낸다.

모형들을 비교할 때 부도를 부도로 바르게 예측한 확률이 같다면 정상기업을 부도기업으로 오판한 확률이 작은 값을 갖는 모형일수록 부도 예측능력이 우월한 모형이라 할 수 있다. 또한 정상을 부도로 예측한 확률이 동일하다면 부도기업을 부도로 예측한 확률이 큰 값을 갖는 모형일수록 부도 예측능력이 우월한 모형이다. 따라서 CAP곡선 아래쪽의 면적이 클수록 부도 예측능력이 우월한 모형이라 할 수 있다. 이를 반영하여 모형의 부도 예측능력을 수치화 한 값이 AR이다. AR은 식 (4)와 같이 이상적인 모형과 무차별 모형 사이에 존재하는 삼각형의 면적을 분모로, 평가모형의 변별력 부분을 분자로 하여 계산한다.

$$AR = \frac{a_R}{a_P} \quad (4)$$

<그림 3>의 오른편에 나타난 ROC 곡선에서 X축은 정상기업을 부도로 예측한 확률을, Y축은 부도기업을 부도로 예측한 확률을 나타낸다. CAP곡선에서와 마찬가지로 무차별 모형은 주어진 모형이 정상과 부도를 구분하는데 전혀 변별력을 제공하지 않는 경우로, 실제 정상기업에 대해 10%를 부도로 예측했다면, 실제 부도기업에 대해서도 10%를 부도로 예측하는 경우가 이에 해당한다. 따라서 원점과 (1,1)점을 지나는 기울기 45°의 직선이 무차별 모형을 나타내는 선이 된다. 여러 모형의 ROC곡선을 비교하는 경우 실제 정상을 부도로 예측하는 비율이 같을 때, 실제 부도기업을 부도로 예측하는 비율이 가장 높은 모형이 성능이 가장 좋은 모형이 될 것이다. AUROC는 모형에 의한 변별력 부분인 ROC곡선 아래부분의 면적을 적분하여 구한 값이다. 변별력이 우수한 모형은 부도위험이 높은 것으로 평가된 낮은 점수에서 정상기업들의 비율이 매우 낮고 부도기업들의 비율은 매우 높게 나타나야 하므로 모형의 변별력이 우수할수록 AUROC는 큰 값을 가질 것이며, 식 (5)와 같이 계산된다.

$$AUROC = \sum_i \left[ \frac{1}{2} Good_i \times Bad_i + (1 - cum\ Good_i) \times Bad_i \right] \quad (5)$$

단,  $Good_i$ 는 신용등급  $i$ 에 속한 정상기업수 / 전체기업수  
 $Bad_i$ 는 신용등급  $i$ 에 속한 부도기업수 / 전체기업수  
 $cum\ Good_i$ 는 낮은 등급에서 높은 등급 순으로 누적된  $Good_i$

AR값은 0과 1사이에 존재하며, AUROC값은 0.5에서 1사이의 값을 갖는다. 두 지표 모두 1에 가까울수록 변별력이 높음을 의미한다. Engelmann,

Hayden and Tasche(2003)이 AR과 AUROC 사이의 관계를 간단한 식으로 증명하였기 때문에 두 가지 방법 중 하나만 사용하더라도 다른 하나는 쉽게 구할 수 있다.

즉, AR과 AUROC의 관계는 다음과 같이 나타난다:  $AR = 2 \times AUROC - 1$ .<sup>17)</sup>

### 3.5.2 Kolmogorov-Smirnov 통계량<sup>18)</sup>

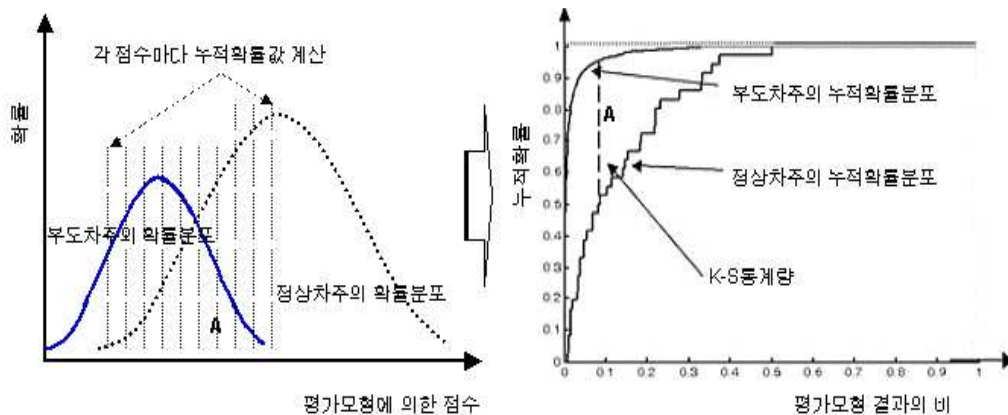
Kolmogorov-Smirnov(K-S) 검정은 정상기업과 부도기업을 구분하여 누적확률분포를 계산하고, 각 등급별로 그 누적확률분포의 차이를 계산하여 변별력을 평가한다. 정상기업은 높은 점수를, 부도기업은 낮은 점수를 갖도록 신용평가모형이 개발되었다고 가정한다면, 정상기업은 높은 점수대에 많은 기업이 분포될 것이고 반대로 부도기업은 낮은 점수대에 많은 기업이 분포될 것이다. K-S 검정은 <그림 4>

의 오른쪽에서 보는 것과 같이 먼저 정상기업과 부도기업의 누적확률분포를 계산하고, 각 평가점수대별로 그 차이를 비교하여 그 중에서 차이가 가장 큰 값을 선택하게 되고, 그것이 K-S 통계량이 된다. 특정 신용평가모형의 변별력이 높은 경우 정상기업과 부도기업은 평가점수에 의해 잘 구분될 것이며, 그에 따라 누적확률분포의 차이가 크게 나타날 것이다.<sup>19)</sup>

## IV. 실증분석 결과와 해석

### 4.1 단일변량분석 결과

T-검증, 단일변량 로짓모형, AR을 사용하여 117개의 재무비율을 대상으로 단일변량 분석을 실시한 결과 금융비용대부채비율, 당좌비율, 이자비용대총



<그림 4> K-S 검정

17) AR과 AUROC 등을 이용한 신용평가모형의 적합성 검정에 대해서는 정완호, 국찬표, 홍광현(2006), “기업 신용도 측정모형의 적합성 비교 연구” 참조.

18) 임종건, 2005, “신용평가시스템에 대한 적합성 검증 방법론 소개”, 금융감독원.

19) 일반적으로 K-S 통계량이 40% 이상일 경우 양호한 수준이라고 할 수 있으며, 20% 미만이면 모형에 대해 재검토 하는 것이 필요하다고 여겨진다.

비용비율, 총부채회전율, 총자본투자효율, 총자산순이익률, 현금비율, 현금성자산비율, 자기자본비율, 차입금의존도 등이 부도예측에 유의한 설명변수로 나타났다.<sup>20)</sup> 이는 외부감사대상기업들의 경우 기업의 성장성이나 수익성 요인보다는, 안정성과 유동성 요인이 기업부실에 보다 중요한 영향을 미침을 말해주는 것으로 외형 성장이나 수익성이 양호하더라도 과도한 이자비용 부담이나 유동성관리에 실패하는 경우 도산위험에 노출됨을 의미한다.

#### 4.2 다변량 로짓분석 결과

다변량 로짓모형의 분석 결과를 이용하여 어떤 재무비율들의 조합이 부도기업과 정상기업을 유의적으로 판별하고 예측하는가를 분석하고 그 결과를 제시한다. 부도예측모형을 구성하기 위해서는 117개의 재무비율 중에서 로짓모형에 사용될 설명변수들을 추출하는 과정이 필요한데, 본 연구에서는 단계적 선택법(stepwise selection method)을 사용하였다. 또 업종별 부도확률의 차이를 감안하여 업종을 더미(dummy)변수로 처리하여 포함하고, 기업규모와 부도확률간의 관련성을 통제하기 위해 총자산에 자연로그(ln)를 취한 값을 변수로 투입하였다.

변수의 선정 전에, 재무비율 자료에 존재하는 극단치와 결측치를 조정하였다. 재무비율을 사용하는 통계분석의 경우 일부 극단치가 전체 결과에 큰 영향을 줄 수 있으며, 어떤 비율의 경우는 결측치가 많이 발생하여 실제 평가모형 운영시 그 의미가 미미해질 수도 있기 때문이다. 이러한 영향을 통제하기 위해서 본 연구에서는 결측치는 해당변수의 중앙값으로 대체하였고, 특정변수에서 극단치는 평균에서

$-2 \times$ 표준편차를 벗어나는 경우 평균 $-2 \times \sigma$ 값으로, 평균에서  $+2 \times$ 표준편차를 벗어나는 경우 평균 $+2 \times \sigma$ 값으로 조정하였다.

분석 결과, 제조업과 건설업 더미변수와 규모변수를 포함한 17개의 변수가 단계적 선택법에 의하여 모형의 설명변수로 선정되었다. 최우추정법에 의한 다변량 로짓분석의 추정결과를 <표 5>에 나타내었다.

모형의 추정결과를 보면 우선 제조업더미와 건설업더미 변수가 부도확률에 유의한 (+)의 영향을 미쳐 제조업과 건설업에 속한 기업들이 다른 업종에 속한 기업들에 비해 부도위험에 상대적으로 크게 노출되어 있음을 보여준다. 이는 <표 3>의 내용과 일관성이 있는 결과이다. 기업규모를 나타내는 총자산 변수는 부도확률에 유의한 음(-)의 영향을 미쳐 기업규모가 작을수록 부도위험에 크게 노출됨을 보여준다.

부채상환능력을 나타내는 금융비용대부채비율은 예상대로 부도확률과 유의한 (+)의 관계를 가진다. 안정성을 나타내는 자기자본비율과 차입금의존도, 활동성을 나타내는 자본금회전율과 총부채회전율, 생산성을 나타내는 총자본투자효율과 유동성을 나타내는 현금성자산비율, 그리고 수익성을 대표하는 영업이익대총자산비율은 모두 예상대로 부도확률에 유의한 음(-)의 영향을 미친다. 그러나 생산성을 나타내는 설비투자효율과 현금흐름변수인 현금흐름대총자본비율, 안정성을 나타내는 유보이익대총자산비율, 수익성을 나타내는 총자산순이익률은 예상과 반대로 부도확률에 (+)의 영향을 미치는 것으로 나타난다. 이는 생산성, 안정성, 현금흐름, 수익성을 나타내는 유사변수들이 모형에 중복되어 포함된 영향이 나타난 것으로 해석된다.

20) 단일변량 분석결과에 대한 자세한 보고는 지면관계상 생략한다.



&lt;표 5&gt; 다변량 로짓회귀분석 추정결과

변수코드	범주	변수	가중치	표준오차	Wald통계량	p-value
Intercept			2.8668	1.6644	2.9666	0.0850
FR010	부채상환능력	금융비용대 부채비율	0.3432	0.0462	55.143	< 0.0001
FR011	비용구조	금융비용대총비용	-0.1061	0.0429	6.1200	0.0134
FR030	생산성	설비투자효율	0.00004	0.0000	5.6308	0.0171
FR042	현금흐름	현금흐름대총자산비율	0.0124	0.0032	14.910	0.0001
FR046	수익성	영업이익대총자산비율	-0.0138	0.0063	4.7516	0.0293
FR065	안정성	유보이익대총자산비율	0.0204	0.0049	16.899	< 0.0001
FR076	안정성	자기자본비율	-0.0288	0.0058	24.298	< 0.0001
FR080	활동성	자본금회전율	-0.0159	0.0059	7.1260	0.0076
FR090	안정성	차입금의존도	-0.0169	0.0036	21.809	< 0.0001
FR092	활동성	총부채회전율	-1.4934	0.2303	42.037	< 0.0001
FR095	생산성	총자본투자효율	-0.0561	0.0089	39.837	< 0.0001
FR096	활동성	총자본회전율	1.4723	0.2607	31.893	< 0.0001
FR097	수익성	총자산순이익률	0.0401	0.0102	15.555	< 0.0001
FR106	유동성	현금성자산비율	-0.0353	0.0079	19.809	< 0.0001
dummy1	산업	제조업	1.4034	0.4006	12.275	0.0005
dummy2	산업	건설업	2.4088	0.4684	26.464	< 0.0001
ln(asset)	규모	총자산	-0.3836	0.0911	17.736	< 0.0001

<표 6>에 모형에 포함된 각 재무비율들을 이용하여 각기 구한 오즈(odds)를 나타내었다.

<표 6>의 추정결과를 보면, 각 변수들이 부도확률에 미치는 개별적인 영향을 파악할 수 있다. 즉, Odds = 부도확률/(1-부도확률)이므로 결국 부도확률 대 정상확률의 비율이다. 따라서 <표 6>의 각 변수들의 Odds는 해당변수가 1단위 증가할 때 부도확률대정상확률의 비율이 몇 배 증가하는 가를 나타낸다. Odds가 1보다 작으면 정상확률에 비해 부도확률이 감소한다는 의미를 나타내고, 1이면 변화없다는 의미를 나타내며, 1보다 크면 정상일 확률보다 부도일 확률이 증가한다는 의미를 나타낸다. 또한 신뢰구간이 1을 포함하고 있으면 통계적 유의성이 없는 것이

고, 1에서 벗어나면 유의성이 있다고 해석할 수 있다. <표 6>에서 부도확률에 의미있는 영향을 미치는 대표적인 변수는 금융비용대부채비율, 총부채회전율, 총자본회전율, 총자산, 그리고 제조업과 건설업 더미변수이다. 금융비용대 부채비율이 한 단위 증가하는 경우 부도확률대정상확률의 비율은 약 1.4배 증가한다. 이는 모집단의 부도확률이 10%라면 금융비용대부채비율이 1% 증가하는 경우 부도확률이 4.44% 증가함을 의미한다. 반면, 총부채회전율의 Odds는 0.225의 값을 갖는다. 이는 총부채회전율이 1% 증가하는 경우 부도확률대정상확률의 비율이 0.225배로 감소함을 나타낸다. 총자산과 유보이익 대총자산비율, 총자산순이익률은 증가시 부도위험이

〈표 6〉 각 재무비율 변수들에 대한 오즈(odds) 추정치

변수코드	범주	변수	Odds	95% Wald 신뢰구간	
FR010	부채상환능력	금융비용대 부채비용	1.409	1.287	1.543
FR011	비용구조	금융비용대총비용	0.899	0.827	0.978
FR030	생산성	설비투자효율	1.000	1.000	1.000
FR042	현금흐름	현금흐름대총자본비용	1.012	1.006	1.019
FR046	수익성	영업이익대총자산비용	0.986	0.974	0.999
FR065	안정성	유보이익대총자산비용	1.021	1.011	1.031
FR076	안정성	자기자본비용	0.972	0.961	0.983
FR080	활동성	자본금회전율	0.984	0.973	0.996
FR090	안정성	차입금의존도	0.983	0.976	0.990
FR092	활동성	총부채회전율	0.225	0.143	0.353
FR095	생산성	총자본투자효율	0.945	0.929	0.962
FR096	활동성	총자본회전율	4.359	2.615	7.266
FR097	수익성	총자산순이익률	1.041	1.020	1.062
FR106	유동성	현금성자산비용	0.965	0.950	0.980
dummy1	산업	제조업	4.069	1.856	8.921
dummy2	산업	건설업	11.132	4.445	27.880
ln(asset)	규모	총자산	0.681	0.570	0.815

감소함을 보여주며, 건설업과 제조업 더미의 Odds는 매우 큰 값을 보여 업종별 특성이 부도위험에 크게 영향을 미침을 보여준다.<sup>21)</sup>

〈표 7〉은 다변량 로짓모형의 분류정확도와 예측력을 나타내는 지표의 추정결과이다. 〈표 7〉의 결과를 보면 현재 모형으로 관측치를 정확히 분류한 비율은 82.3%이며, 12.3%는 관측치와 다른 결과가 나왔다. 모형의 변별력을 나타내는 AUROC는 85.0%, AR값은 70.0%를 보여 구축된 모형이 상당히 높은 예측력을 지닌 모형임을 보여준다.

〈표 7〉 다변량 로짓모형의 분류정확도와 예측력 지표

Concordant(%)	82.3
Discordant(%)	12.3
Tied(%)	5.4
Pairs	5,026,910
AR(%)	70.0
AUROC(%)	85.0

〈표 5〉의 결과를 이용하면 부도예측을 위한 로짓함수는 다음과 같이 구성된다.

21) 총자본회전율의 Odds는 4.359로 총자본회전율이 1단위 증가할 때 부도위험이 크게 증가함을 나타낸다. 이는 〈표 5〉에서와 마찬가지로 활동성을 나타내는 유사변수들이 포함된 영향이 반영되어 나타난 결과로 보인다.

$$y = 2.8668 + 0.3432FR010 - 0.1061FR011 \\ + 0.000041FR030 + 0.0124FR042 - 0.0138FR046 \\ + 0.0204FR065 - 0.0288FR076 - 0.0159FR080 \\ - 0.0169FR090 - 1.4934FR092 - 0.0561FR095 \\ + 1.4723FR096 + 0.0401FR097 - 0.0353FR105 \\ + 1.4034dummy1 + 2.4098dummy2 \\ - 0.3836\ln(asset)$$

위의 로짓함수를 이용하여 예상부도확률을 구할 수 있다. 즉, 위의 식으로부터 도출되는  $y$ 값을 다시 지수함수(exponential function)를 이용하여 예상 부도율 값을 다음과 같이 얻는다:

$$E(y) = \frac{\exp(z)}{1 + \exp(z)} \quad \text{or} \quad \frac{1}{1 + \exp(-z)}$$

#### 4.3 예상부도확률을 이용한 등급화와 모형의 성능

구성된 로짓모형의 실제적용 가능성을 보기 위해 표본기업에 대해 추정된 로짓함수를 이용하여 예상 부도확률을 구하고 그 크기에 따라 전체표본을 9개의 그룹으로 구분하였다. 이후 각 그룹에 속한 기업들의 실제 부도자료를 이용하여 그룹별 부도기업과 건전기업, 그리고 부도율을 구하여 <표 8>에 나타내었다. <그림 5>는 <표 8>의 결과를 나타낸 것이다.

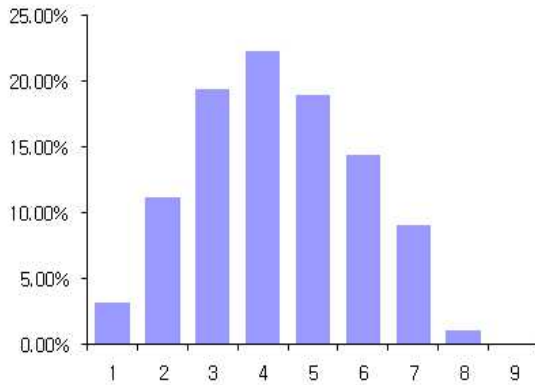
등급 구분은 금융감독원(2005)에서 제시하는 기준인 “신용리스크 내부등급 기본 세부지침”을 참조하였다.<sup>22)</sup> 이에 따르면 등급을 10개로 구분하는 경우

<표 8> 추정된 로짓함수를 이용하여 구한 예상부도확률에 따른 등급구분과 등급별 부도율

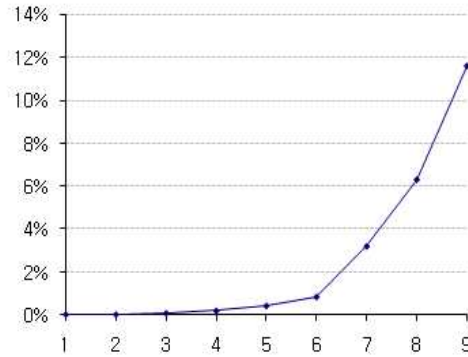
예상부도확률(%)		등급	표본기업	구성비율	실제 부도 또는 정상 기업 분포		
From	To				부도기업	정상기업	부도율
0.0000	0.0001	1	1,665	3.20%	-	1,665	0.00%
0.0001	0.0100	2	5,839	11.22%	-	5,839	0.00%
0.0100	0.1000	3	10,144	19.49%	6	10,138	0.06%
0.1000	0.3000	4	11,609	22.30%	20	11,589	0.17%
0.3000	0.7000	5	9,883	18.99%	40	9,843	0.40%
0.7000	1.5000	6	7,541	14.49%	62	7,479	0.82%
1.5000	5.0000	7	4,746	9.12%	151	4,595	3.18%
5.0000	15.000	8	556	1.07%	35	521	6.29%
15.000	100.00	9	69	0.13%	8	61	11.59%
		합계	52,052	100.00%	322	51,370	0.62%

22) 주요내용은 다음과 같다. (1) 차주등급은 부도가능성에 근거하여 차주를 서열화하여야 한다. 예를 들어, 은행이 10등급 체계를 사용하는 경우 1등급이 가장 신용도가 높은 등급이 되고 10등급이 부도등급을 나타낸다고 하면 그 사이의 2~9등급은 단계적으로 부도가능성이 증가해야 함을 의미한다. (2) 차주등급의 수와 관련하여 부도나지 않은 차주와 부도차주에 대해서 각각 최소 7개 이상, 최소 1개 이상의 등급이 있어야 하며, 차주등급의 수와 등급 간 차주 분포의 적정성을 충분히 설명할 수 있어야 한다. (3) 은행에서 사용하는 신용평가시스템 등급의 수는 적용 대상 포트폴리오의 차주 위험을 의미 있게 차별화 할 수 있다고 판단할 수 있을 만큼 충분해야 한다. (4) 부도 아닌 차주에 대해서 최소 7개 이상의 등급으로 세분화하도록 하고 있으며, 특정 등급에 집중되지 않도록 한다. (5) 차주들이 특정 등급에 집중된 경우에는 모형 개발이 은행의 모형 운영 대상(목표 시장)과 일관되지 않거나 의도했던 대로 절차가 수행되지 않은 근거(절차의 불안정성 또는 신용등급평가 철학의 비일관성 등)가 된다. (6) 차주 또는 익스포저가 특정 등급에 집중되어 있고 차주 간 재무상태의 편차가 큰 경우에는 차주 위험을 의미 있게 차별화했다고 보기 어렵다. 한 등급에 30% 이상 또는 3개의 등급에 80% 이상 유지되고 있는 경우에는 자체 질문을 통해서 적정성 여부를 분석하고 문서화해야 한다.

등급별 분포



등급별 부도율



〈그림 5〉 등급별 차주 분포와 부도율의 형태

10등급은 실제 부도발생 기업에 부여하는 등급이기 때문에 모형에서는 10등급을 산출하지 않으며 9등급까지 부여한다. 등급 계량화 과정을 거쳐 등급화된 모형의 등급분포와 부도율은 〈표 8〉과 같다.<sup>23)</sup>

〈표 8〉의 결과를 보면 9개 등급에서 등급이 낮아질수록 부도율이 증가하는 우상향의 지수함수 형태를 보이고 있다. 이는 단계적으로 부도가능성이 증가한다는 것을 의미한다. 차주가 특정 등급에 집중되어 있지 않고 등급 간 차주수가 적절히 분포되어 있다. 한 등급에 가장 많은 차주수가 분포하는 것이 22%이며 3개의 등급에 가장 많은 차주가 분포하는 경우가 약 60%이다. 등급 간 분포는 상위등급으로 약간 치우친 것을 볼 수 있으며, 각 등급의 실제 부도율은 역전 현상이 나타나지 않는다. 〈표 8〉과 〈그림 5〉는 본 연구에서 구축한 로짓모형을 이용하여 구한 예상부도확률이 각 등급별 실제 부도율의 변화 행태를 잘 설명하며, 실제 기업의 부도예측에 유용

하게 이용될 수 있음을 보여주는 것이다.

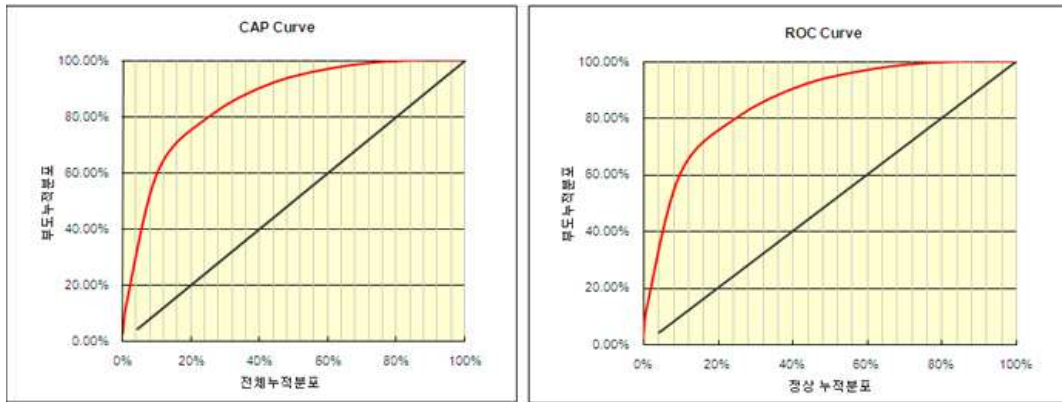
등급 계량화 과정을 거쳐 등급화된 모형의 변별력을 살펴보기 위해 AUROC값과 AR값을 〈표 9〉에 K-S 통계량을 〈표 10〉에 나타내었다. 〈그림 6〉과 〈그림 7〉은 이를 그림으로 나타낸 것이다. 〈표 10〉과 〈그림 8〉에는 등급화된 모형의 K-S 통계량을 나타내었다

〈표 9〉의 결과는 모형의 변별력 검증 결과 AUROC 값은 84.97%, AR 값은 69.95%로 상당히 높은 변별력을 가짐을 보여준다. 또 〈표 10〉의 결과는 모형의 K-S 통계량 또한 55.04%로 상당히 높은 수준을 가짐을 보인다. 이러한 결과들은 본 연구에서 구축된 다변량 로짓분석을 이용한 부도예측모형이 한국시장의 외부감사대상기업들의 부도예측에 유용하게 이용될 수 있음을 보여주는 것이다.

23) 산업(업종)별, 연도별로도 상세한 분석을 하였으나 그에 대한 상세한 보고는 생략한다.

〈표 9〉 등급화된 모형의 성능(AUROC)

등급	표본기업수			등급분포(비율, %)			누적분포(비율, %)			AUROC
	부도	정상	합계	부도	정상	전체	부도	정상	전체	
1	-	1,665	1,665	0.00	3.22	3.20	100.00	100.00	100.00	0.00
2	-	5,839	5,839	0.00	11.29	11.22	100.00	96.78	96.80	0.00
3	6	10,138	10,144	1.86	19.60	19.49	100.00	85.49	85.58	0.45
4	20	11,589	11,609	6.21	22.40	22.30	98.14	65.90	66.10	2.81
5	40	9,843	9,883	12.42	19.03	18.99	91.93	43.49	43.79	8.20
6	62	7,479	7,541	19.25	14.46	14.49	79.50	24.47	24.81	15.94
7	151	4,595	4,746	46.89	8.88	9.12	60.25	10.01	10.32	44.28
8	35	521	556	10.87	1.01	1.07	13.35	1.13	1.20	10.80
9	8	61	69	2.48	0.12	0.13	2.48	0.12	0.13	2.48
합계	322	51,370	52,052	100	100	100	AUROC=84.97% AR=69.95%			



〈그림 7〉 등급화된 모형의 CAP곡선과 ROC곡선

#### 4.4 강건성분석

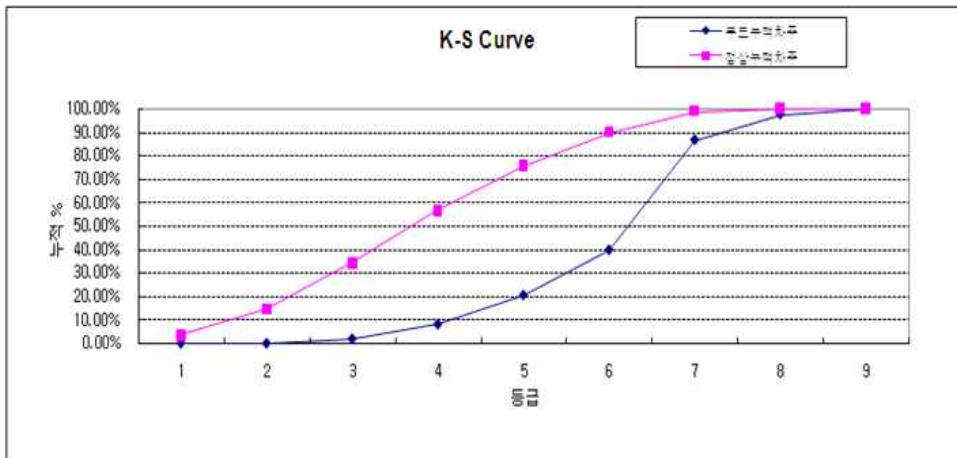
이 절에서는 앞서 분석된 결과의 강건성을 확인하기 위해 추가적인 검증을 실시한다. 우선 표본에 존재하는 극단치와 결측치가 모형의 결과에 어떤 영향을 주는지를 확인하기 위해 극단치와 결측치를 조정

하지 않은 데이터에 기초해 모형을 구축하고 그 결과를 확인한다.

극단치와 결측치를 조정하지 않은 데이터에 기초해 구성한 모형의 Power 지표는 다음의 〈표 11〉과 같다. 로짓함수를 구성하는 일부 재무비율이 같은 범주내의 유사 재무비율로 대체되는 변화가 있으나,

〈표 10〉 등급화된 모형의 K-S 통계량

등급	표본기업수				등급분포비율(%)		누적분포비율(%)		K-S
	부도	정상	합계	부도율	부도	정상	부도	정상	
1	-	1,665	1,665	0.00%	0.00	3.22	0.00	3.22	3.22
2	-	5,839	5,839	0.00%	0.00	11.29	0.00	14.51	14.51
3	6	10,138	10,144	0.06%	1.86	19.60	1.86	34.10	32.24
4	20	11,589	11,609	0.17%	6.21	22.40	8.07	56.51	48.43
5	40	9,843	9,883	0.40%	12.42	19.03	20.50	75.53	<b>55.04</b>
6	62	7,479	7,541	0.82%	19.25	14.46	39.75	89.99	50.24
7	151	4,595	4,746	3.18%	46.89	8.88	86.65	98.87	12.23
8	35	521	556	6.29%	10.87	1.01	97.52	99.88	2.37
9	8	61	69	11.59%	2.48	0.12	100.00	100.00	0.00
합계	322	51,370	52,052	0.62%	100.00	100.00			



〈그림 8〉 등급화된 모형의 성능(K-S 통계량)

전체 모형에 큰 변화는 없으며 각 변수가 미치는 영향 역시 큰 차이는 없다. 〈표 11〉을 보면 모형이 관측치를 정확히 예측한 비율은 86.6%이며, 10.4%는 관측치와 다른 결과가 나왔다. AUROC는 88.1%, AR값은 76.2%를 보여 극단치와 결측치를 조정하

기 전의 모형에 비해 더 높은 변별력을 보여준다. 〈표 11〉의 결과는 앞서 얻어진 분석결과가 자료에 존재하는 극단치나 결측치의 영향으로 인한 결과가 아님을 보여주는 것이다.

〈표 11〉 다변량 로짓모형의 분류정확도와 예측력 지표:  
극단치와 결측치를 조정하지 않은 경우

Concordant(%)	86.6
Discordant(%)	10.4
Tied(%)	3.0
Pairs	5,026,910
AR(%)	76.2
AUROC(%)	88.1

다음으로 계절적인 영향과 산업별 특성을 살펴보기 위해 앞서 구축된 모형을 이용하여 결산연도 재무제표 기준으로 2003, 2004, 2005, 2006년의 연도별 분석과, 제조업, 건설업, 도소매업, 서비스업의 산업별 분석을 실시하였다.

〈표 12〉는 연도별 분석결과 추정된 AR값과 K-S 통계량을 나타내며, 〈표 13〉은 산업별 분석결과 추정된 모형의 AR값과 K-S통계량을 보여준다.

구축된 모형의 성능에 대한 시계열분석 결과인 〈표 12〉를 보면, 연도별로 모형의 변별력을 나타내는 AR값은 약 64%에서 74%, K-S 통계량은 약 49%에서 62%의 분포를 가져 상당히 높은 수준을 보여준다. 2004년과 2005년은 2003년과 2006년에 비해 더욱 높은 변별력을 나타내는데 이는 전자가 후

자에 비해 더욱 많은 부도 차주수를 지니고 있으며, 부도율 또한 월등히 높는데 원인을 찾을 수 있을 것이다. 일반적으로 전체 차주에 비해 부도 차주가 많을수록 모형의 변별력은 증가한다. 모형의 연도별 변별력 분석결과는 안정적인 것으로 나타내는데, 이는 모형이 시계열적으로 강건하다는 것을 나타낸다. 산업별 분석결과인 〈표 13〉을 보면 산업별로 상당한 차이를 보여준다. 도소매업과 제조업의 변별력이 상대적으로 우수하며 건설업과 서비스업의 변별력이 떨어진다. 특히 건설업종의 변별력이 다른 산업에 비해 떨어진다. 건설업의 AR값은 52.5%이며 K-S 통계량은 40.9%이다. 반면 도소매업의 AR값은 71.6%, K-S 통계량은 61.4%를 보여 상대적으로 높은 수준의 변별력을 보여준다. 이러한 결과는 한국시장의 외부감사대상기업들의 부실예측모형을 구축시 산업별 특성을 반영하는 것이 중요함을 보여주는 것이다.

## V. 결 론

본 연구는 상장기업들을 대상으로 이루어진 선행 연구와 달리 우리나라의 외부감사대상기업 전체를

〈표 12〉 모형의 연도별 변별력 분석 결과(2003년~2006년)

결산연도	2003	2004	2005	2006	전체
AR	65.04%	74.22%	72.91%	63.61%	70.0%
K-S	53.39%	62.21%	56.18%	48.54%	55.04%

〈표 13〉 모형의 업종별 변별력 분석 결과

구분	제조업	건설업	도소매업	서비스업	전체
AR	72.8%	52.5%	71.6%	54.1%	70.0%
K-S	58.2%	40.9%	61.4%	39.4%	55.04%

대상으로 재무비율을 이용한 부도예측 모형을 구축하고 그 적용가능성을 살펴보았다. 또 금융결제원의 당좌거래정보를 이용하여 부도기업을 포괄적으로 선별함으로써 보다 풍부한 데이터를 실증분석에 사용하였으며, 기업의 부실을 설명해줄 수 있는 변수를 파악하기 위해 구성가능한 117개의 대량의 재무비율을 사용하여 보다 실무적인 방법을 사용한 점 역시 기존연구와 차별화되는 점이다. 모형의 변별력을 파악하기 위해 AUROC, AR, K-S 통계량 등을 사용한 것도 본 연구의 특징이다.

결산연도 기준으로 2003년~2006년의 외부감사대상기업들의 데이터를 이용하여 분석한 본 연구의 주요결과는 다음과 같다. 단일변량 분석결과 금융비용 대 부채비용, 당좌비용, 이자비용 대 총비용, 총 부채회전율, 총자본투자효율, 총자산순이익률, 현금비용, 현금성자산비용, 자기자본비용, 차입금의존도 등이 기업부실을 예측하는 데 유의한 변수임을 보여준다. 이는 외부감사대상기업들의 경우 기업의 성장성이나 수익성 요인보다는, 안정성과 유동성 요인이 기업부실에 보다 중요한 영향을 미침을 말해주는 것으로 외형 성장이나 수익성이 양호하더라도 과도한 이자비용 부담이나 유동성관리에 실패하는 경우 도 산위험 노출됨을 의미한다.

다변량 로짓모형을 이용하여 분석한 결과는 금융비용대부채비용, 자기자본비용, 차입금의존도, 현금성자산비용, 총부채회전율, 영업이익대총자산비용, 총자본투자비용, 기업규모, 건설업 더미, 제조업더미 등 17개의 변수가 부실예측모형을 구성하는 변수로 선택되었다. 모형의 추정결과 부실예측의 정확도(concordant ratio)는 82%로 나타났으며, 모형의 변별력을 나타내는 지표인 AUROC값은 85%, K-S 통계량은 55%로 매우 양호하게 나타났다. 이러한 결과는 데이터에 존재하는 극단치와 결측치에

영향을 받지 않는 결과이며, 매 연도별 분석에서도 큰 차이를 보이지 않아 시계열적으로도 강건한 결과이다. 이는 본 연구에서 구축한 부도예측모형이 한국시장의 외부감사대상기업의 부도예측에 유용하게 사용될 수 있음을 의미한다. 산업별로 분석한 결과는 모형의 예측력이 산업별로 상당한 차이가 있음을 보여주어 재무비율을 이용한 부도예측모형의 구축에 산업별 특성이 적절하게 반영될 필요가 있음을 보여준다.

본 연구에서는 외부감사대상기업 전체에 적용가능한 모형의 구축에 초점을 맞추어 재무비율만을 이용한 가능한 정보로 사용하였으나, 주가 등 시장정보와 질적인 요인들을 다양하게 반영하는 연구로 확대될 필요가 있다. 금융결제원의 당좌거래정보에 기초하여 부도기업을 파악하였으나 은행연합회의 자료를 이용한 연구는 모형의 질적 개선에 큰 도움을 줄 수 있을 것이다. 또한 본 연구에서 보인 산업별 특성의 차이를 반영한 부도예측모형의 구축과 평가는 실무적으로 의미있는 결과를 제공해줄 수 있을 것이다.

## 참고문헌

- 강종만, 홍성의(1999), “부실예측모형의 적합성 분석,” **증권금융연구**, 제5권 1호, 83~110.
- 박정운(2000), “재무정책과 기업부실예측,” **재무관리논총**, 제6권, 93~116.
- 송인만(1987), “기업부실예측모델의 재정립을 통한 기업부실원인과의 연계에 대한 실증적 연구,” **한국경제연구**, 제15권, 113~146.
- 안성만(2009), “기업의 부도예측에 관한 실증연구,” 서울시립대학교 석사학위논문
- 이건창(1993), “기업도산예측을 위한 통계적 모형과 인공



- 지능 모형간의 예측력 비교에 관한 연구: MDA, 귀납적 학습방법, 인공신경망,” **한국경영과학회지**, 제18권 제2호, 57~81.
- 이계원(1993), “회계정보에 의한 기업부실예측과 시장반응,” **회계학연구** 제16호, 49~77.
- 임종건(2005), “신용평가시스템에 대한 적합성 검증 방법론 소개,” 금융감독원.
- 장휘용(1998), “비금융 상장기업의 부실예측모형,” **재무관리연구**, 제15권 1호, 299~327.
- 전춘옥, 1984, 기업도산론, 무역경영사.
- 정완호, 국찬표, 홍광현(2006), “기업 신용도 측정모형의 적합도 비교 연구,” **금융학회지**, 제11권 제2호, 67-104.
- 금융감독원 신BIS실(2007), “위험가중자산에 대한 자기자본비를 산출기준”.
- 금융감독위원회(2004), “신용리스크 내부등급법 기본 세부 지침(안)”.
- 금융감독위원회(2005), “신용리스크 내부등급법 기본 세부 지침(안)”.
- 금융결제원 당좌거래정보, <http://www.kftc.or.kr/kftc/info/dishonor/period.jsp>.
- Altman, E.I.(1968), “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *Journal of Finance* 23, 589-609.
- Altman, E.I., R. Haldeman and P. Narayanan (1977), “ZETA Analysis : A New Model to Identify Bankruptcy Risk of Corporations,” *Journal of Banking and Finance*, 29-54.
- Altman, E.I.(1983), *Corporate Financial Distress: A Complete Guide to predicting, Avoiding and Dealing with Bankruptcy*, New York, John Wiley & Sons, 120-124.
- Altman, E.I.(1996), “Distress Classification of Korean Firm,” in Altman E. ed., *Bank Credit Risk Management*, *Korea Institute of Finance*, 381-410.
- Beaver, W.(1966), “Financial Ratios as Predictors of Failure,” *Journal of Accounting Research*, 71-111.
- Ohlson, J.A.(1980), “Financial Ratios The Probabilistic of Bankruptcy,” *Journal of Accounting Research*, 109-131.
- Stein, R.M.(2002), “Benchmarking Default Prediction Models: Technical Report #020305.” Moody’s KMV,
- Zavgren, C.V.(1985), “Assessing the Vulnerability to Failure of American Industrial Firm : A Logistic Analysis,” *Journal of Business Finance and Accounting* 12(1), 19-45.
- Zmijewski, M.E.(1984), “Methodological Issues Related to the Estimation of Financial Distress Prediction Models,” *Journal of Accounting Research* 22 Supplement, 59-82.

## 〈부록〉 데이터 정제과정

본 연구에서 사용된 데이터 정제과정을 정리하면 다음과 같다.

### Step 1. 재무정보 정제

본 연구에서 사용한 재무제표는 한국신용평가정보(KIS)에서 제공하는 2001년부터 2006년(결산연도 기준)까지의 자료이다. 재무제표는 기본적으로 하위 계정과목들의 합계가 상위 계정과목들의 합계와 일치해야 하며, 재무상태표(대차대조표)의 경우에는 대변과 차변이 반드시 일치해야만 한다. 한국신용평가정보의 재무제표는 천원 단위로 되어 있기 때문에 반올림 등으로 인한 대차가 일치하지 않는 오류가 다수 존재한다. 따라서 대차가 일치하지 않거나 하위 계정과목의 합계가 상위 계정과목의 합계

와 일치하지 않는 재무제표는 모두 삭제하였다.<sup>24)</sup> 구체적인 정제규칙은 다음의 〈표 A1〉과 같다.

원시 데이터로부터 재무제표의 계정과목들과 관련한 정제규칙을 적용시켜 1차적으로 불량한 재무제표 데이터를 삭제한 결과 삭제된 데이터는 5,629건으로 전체 재무제표 데이터 905,480건의 0.62% 정도이다. 다음으로는 결산일자 변경 등에 의해 회계기간이 1년 미만인 재무정보는 삭제하였다. 이 과정에서 2,067건(0.23%)의 데이터가 삭제되었다. 재무비율을 생성하기 위해서는 최소한 연속 2개년 이상의 결산 재무제표가 필요하다. 따라서 연속 2개년 재무제표가 존재하지 않는 경우는 삭제하였다. 한편, 부도관측기간과의 일치를 위해서 결산 월이 12월이 아닌 데이터는 삭제하였다. 이 과정에서 삭제된 데이터는 모두 22,121건으로 약 2.5%이다. 이상의 재무제표 정제과정을 나타내면 〈그림 A1〉과 같다.



〈그림 A1〉 재무정보 정제과정

24) 중요성의 기준으로 보아 만원 단위까지의 오차는 허용하였다.

〈표 A1〉 재무제표 정제규칙과 정제건수

Num	Cleansing Rule	건수	비율
C01	BS의 대자가 일치하지 않는 경우 삭제(자산총계와 부채와자본총계가 서로 일치하지 않는 경우)	1	0.02%
C02	부채와자본총계가 부채총계, 자본총계의 합계와 일치하지 않는 경우 삭제	22	0.39%
C03	자산총계가 세부항목(유동자산, 비유동자산, 임대주력자산) 합계와 일치하지 않는 경우 삭제	22	0.39%
C04	부채총계가 세부항목(유동부채, 비유동부채, 이연부채) 합계와 일치하지 않는 경우 삭제	10	0.18%
C05	자본총계가 세부항목(자본금, 자본잉여금, 이익잉여금, 자본조정, 기타포괄손익누계액) 합계와 일치하지 않는 경우 삭제	27	0.48%
C06	유동자산이 세부항목(당장자산, 채고자산) 합계와 일치하지 않는 경우 삭제	10	0.18%
C07	비유동자산이 세부항목(투자자산, 유형자산, 무형자산, 기타비유동자산, 이연자산) 합계와 일치하지 않는 경우 삭제	22	0.39%
C08	유동부채 세부항목의 합계가 유동부채와 일치하지 않는 경우 삭제	35	0.62%
C09	비유동부채 세부항목의 합계가 비유동부채와 일치하지 않는 경우 삭제	238	4.23%
C10	자본잉여금과 세부항목(자본준비금, 파정가적립금, 기타) 합계가 일치하지 않는 경우 삭제	-	0.00%
C11	자본조정과 세부항목 합계가 일치하지 않는 경우 삭제	1,439	25.56%
C12	당장자산 세부항목의 합계가 당장자산과 일치하지 않는 경우 삭제	35	0.62%
C13	채고자산 세부항목의 합계가 채고자산과 일치하지 않는 경우 삭제	-	0.00%
C14	유형자산 세부항목의 합계가 유형자산과 일치하지 않는 경우 삭제	15	0.27%
C15	무형자산 세부항목의 합계가 무형자산과 일치하지 않는 경우 삭제	4	0.07%
C16	투자자산 세부항목의 합계가 투자자산과 일치하지 않는 경우 삭제	8	0.14%
C17	판관비 세부항목의 합계가 판관비와 일치하지 않는 경우 삭제	-	0.00%
C18	영업외수익 세부항목의 합계가 영업외수익과 일치하지 않는 경우 삭제	65	1.15%
C19	영업외비용 세부항목의 합계가 영업외비용과 일치하지 않는 경우 삭제	71	1.26%
C20	당기총재조원가가 세부항목(원재료, 노무비, 경비) 합계와 일치하지 않는 경우 삭제	7	0.12%
C21	경비가 세부항목 합계와 일치하지 않는 경우 삭제	36	0.64%
C22	이익잉여금차분액이 세부항목 합계와 일치하지 않는 경우 삭제	2	0.04%
C23	결손금차분액이 세부항목 합계와 일치하지 않는 경우 삭제	-	0.00%
C24	자산총계가 결측값이거나 0보다 작거나 같은 경우 삭제	164	2.91%
C25	유동자산<0 인 경우 삭제	133	2.36%
C26	비유동자산<0 인 경우 삭제	15	0.27%
C27	유동부채<0 인 경우 삭제	141	2.50%
C28	비유동부채<0 인 경우 삭제	443	7.87%
C29	당기차입금<0 인 경우 삭제	32	0.57%
C30	장기차입금<0 인 경우 삭제	8	0.14%
C31	유동상장기부채<0 인 경우 삭제	-	0.00%
C32	장기상채<0 인 경우 삭제	4	0.07%
C33	매출액=0 & 영업이익=0 & 당기순이익>0 인 경우 삭제	-	0.00%
C34	매출액, 영업이익, 당기순이익항목의 값산이 0이거나 결측값인 경우 삭제	1,138	20.22%
C35	매출액<0 인 경우 삭제	46	0.82%
C36	매출원가<0 인 경우 삭제	72	1.28%
C37	판관비<0 인 경우 삭제	7	0.12%
C38	이자비용+상채<0 인 경우 삭제	7	0.12%
C39	당기순이익이 결측값인 경우 삭제	2,671	47.45%
재무제표 Cleansing 결과 삭제되는 데이터 건수		5,629	100%

## Step 2. 재무비율 생성 &amp; 기업개요정보 결합

두 번째 정제과정은 재무제표정보 정제과정을 거쳐 생성된 재무비율정보와 기업개요정보를 결합하는 과정이다. 그 전에 기업개요정보 중 법인번호, 사업자번호, 산업분류코드가 잘못 기입된 경우의 데이터

를 삭제하였다. 또한 업체개요정보에는 법인번호, 사업자번호, 산업분류코드, 회사설립일/창립일, 등에 관한 정보를 지니고 있는데, 이는 각 업체마다 유일하게 존재하는 정보이다. 즉, 법인번호 또는 사업자번호가 유일해야 한다. 따라서 법인번호 또는 사업자번호가 중복된 개요정보는 모두 삭제하였다.<sup>25)</sup>

25) 만약 특정 법인번호가 중복되는 것이 두 개 존재한다면 어느 것이 오류인지를 하나씩 확인하기가 쉽지 않기 때문이다.

이렇게 해서 삭제된 데이터 건수는 2,476건으로 업체 개요정보 전체의 0.3%정도에 해당한다. 업체개요와 재무정보가 결합되는 과정에서 삭제되는 데이터 건수는 10,243건으로 재무제표정보 데이터 875,663건의 1.17%에 해당한다. 이상의 과정을 그림으로 나타내면 <그림 A2>와 같다.

### Step 3. 부도정보 정제

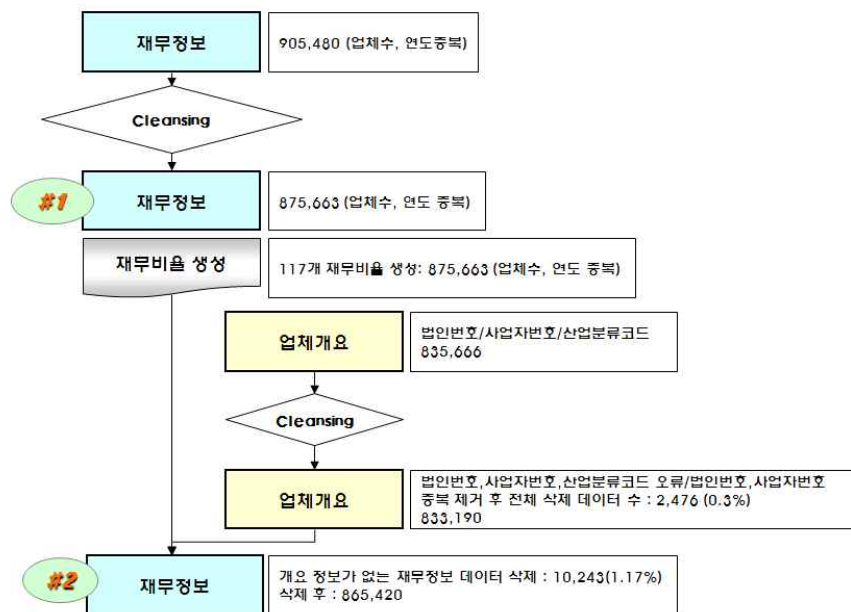
본 연구에서 사용한 금융결제원의 부도정보는 2005년 1월 1일부터 2007년 12월 31일까지 발생한 당좌수표정지 또는 약속어음부도로 인해 당좌거래가 정지된 기업이다. 전체 부도 차주는 개인사업자와 법인을 모두 포함하고 있는데, 그 중 법인만을 선별하고 앞서 정의한 부도 인식 기준에 따라 정제한 후의 데이터는 3,534개 업체이다.

### Step 4. 재무정보와 부도정보의 결합

부도정보와 재무정보가 결합되는 #4 과정은 앞에서 살펴본 부도 인식 기준(방법2)에 따라 최초 부도시점을 부도로 인식하며, 그 이후에는 회복된다 하더라도 부도 차주로 간주하고 모집단에서 삭제하였다. 이에 대한 상세한 과정은 <그림 A4>와 <그림 A5>에 나타나 있다. 앞서의 정제과정을 거친 재무정보와 부도정보를 결합한 결과 838,432건의 데이터가 생성되었다.

### Step 5. 감사의견 정보 결합 & 기타 정제 과정

마지막 정제과정은 최종적으로 연구 모집단을 추출해 내는 과정인데, 감사의견 정보를 결합하기 전에 회사설립일 이전의 결산 일자를 지닌 업체는 제



<그림 A2> 재무정보와 업체개요정보의 결합 과정



〈그림 A3〉 부도정보 정제

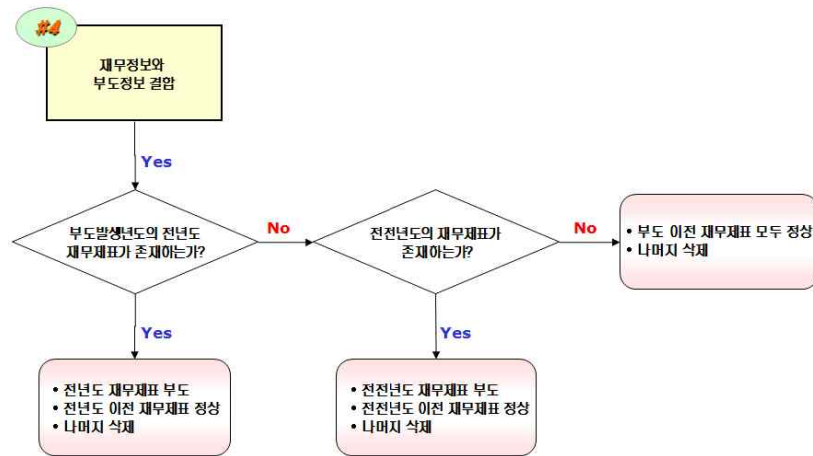


〈그림 A4〉 재무정보와 부도정보의 결합 과정 1

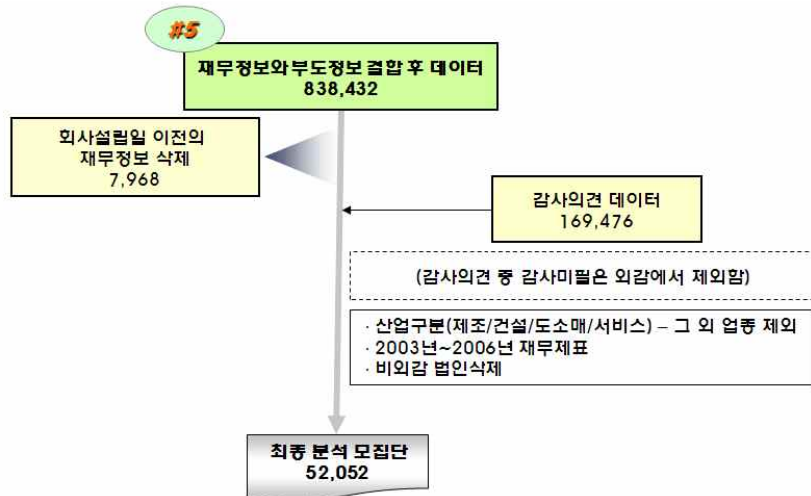
외시켰다.<sup>26)</sup> 감사의견정보 결합 후 제조/건설/도소매/서비스업에 해당하지 않는 기업들은 삭제하였고 2003년 ~ 2006년 12월 결산 법인들만을 선별하였다. 또한 비외감 법인들은 삭제하였다. 즉, 외부감

사의견이 존재하는 업체들만 선별하였다. 외부감사 대상기업으로 최종 분석집단으로 총 기업수는 52,052 개이며 선정된 부도기업 수는 322개이다. 이 과정은 〈그림 A6〉에 나타나 있다.

26) 불확실한 재무제표 결산일자로 판단해서 삭제하였다.



〈그림 A5〉 재무정보와 부도정보의 결합 과정 2



〈그림 A6〉 감사의견 결합과 기타 정제과정

#### Step 6. 분석 대상 표본 구성

6단계의 정제과정을 거쳐 최종적으로 구성된 표본은 제조/건설/도소매/서비스업에 포함되는 외부감사대상 기업 중 2003년 ~ 2006년의 결산연도 재무제표를 이용할 수 있는 12월 말 결산법인들이다.

표본을 구성하는 총 기업-연도 수는 정상기업과 부도기업을 포함해 모두 52,052개로 대부분의 외감기업을 포함하고 있으며, 이중 부도기업은 322개로 한국은행에서 부도로 정의하는 2005년 1월 1일부터 2007년 12월 31일까지의 부도기업을 대부분 포함하고 있다.

## Corporate Bankruptcy Prediction Using Financial Ratios: Focused on the Korean Manufacturing Companies Audited by External Auditors\*

Jong Won Park\*\* · Sung Man Ahn\*\*\*

### Abstract

After the 1997 Asian financial crisis South Korea experienced a serious financial hardship. Immediate economic effects of this crisis were felt in all areas of the Korean business world, and many companies were filed for bankruptcy. The main purpose of this study is to develop a bankruptcy prediction model and to examine the power and accuracy ratio of the model using the data of Korean companies. We use all manufacturing companies audited by external auditors as sample firms during the period from 2003 to 2007 and use the transaction information of the current bank account from the KFTCI(Korea Financial Telecommunications & Clearings Institute) to identify the right time for bankruptcy. Multi-variate Logistic Regression was used to form a prediction model with 117 financial ratios. The major findings of this study are as follows. First, we obtained 17 financial ratios from the stepwise selection method of logistic model and used these ratios in our bankruptcy prediction model based on multi-variate analysis as key input variables. The estimation results show that the concordant ratio which represents the discriminant power of the model is 82%. Second, The predictive power of the bankruptcy prediction model, AUROC is 85% and Kolmogorov-Smirnov statistics is 55.04%. These results mean that the model developed in this study can be used as the bankruptcy prediction model for external-audit-companies in the Korean markets.

Key words: Bankruptcy Prediction, Companies Audited by External Auditors, Financial Ratio, Multi-variate Logistic Regression, AUROC, Kolmogorov-Smirnov(K-S) Statistic.

---

\* This work was supported by the 2013 Research Fund of the University of Seoul.

\*\* College of Business Administration, University of Seoul(parkjw@uos.ac.kr), Corresponding Author,

\*\*\* NH Bank(asm1230@nonghyup.com)