

유연한 분포함수를 이용한 기업도산예측 모형

최필선

건국대학교 상경대학 국제무역학과 조교수
(pchoi@konkuk.ac.kr)

차미현

경희대학교 경영대학 회계학부 박사과정
(colorbase@hanmail.net)

민인식(교신저자)

경희대학교 정경대학 경제학부 조교수
(imin@khu.ac.kr)

기업 도산에 관한 연구는 학술적으로뿐만 아니라 현실적으로 매우 중요한 문제이다. 이러한 기업 도산을 예측하기 위해 오래전부터 다양한 실증연구가 진행되어 왔지만 추정방법론, 설명변수의 선택, 도산기업과 생존기업의 표본선택 문제 등에 따라 다양한 결과가 나타나고 있다. 본 연구에서는 이항종속변수 회귀모형에서 기존의 정규분포나 로지스틱분포 등 극히 제약적인 분포함수가 아니라 이보다 훨씬 유연한 분포인 S_U -정규분포를 오차항으로 가정함으로써 추정치의 일치성을 담보할 수 있는 모형을 제안한다. 몬테카를로 시뮬레이션 결과 우리가 제안한 S_U -정규분포 모형이 기존의 로짓 모형에 비해 오차항의 비대칭성을 포착하는데 있어 더 우월함을 보였다. 또한 2000년~2005년 기간의 상장기업을 대상으로 도산예측 모형을 추정한 결과, 도산여부를 분류하는 판별정확도와 예측정확도 측면에서 S_U -정규분포 모형이 로짓 모형에 비해 더 우월한 것으로 나타났다. 설명변수가 도산확률에 미치는 한계효과에 있어서도 S_U -정규분포 모형과 기존의 로짓 모형의 결과 사이에 큰 차이가 있는 것으로 나타났다. 이러한 실증분석 결과는 본 연구에서 제시하는 모형을 이용하여 기업의 도산에 영향을 미치는 요인과 그 파급효과를 보다 정확하게 계량화할 수 있다는 점에서 의의가 있다고 할 수 있다.

주제어: S_U -정규분포; 도산예측모형; 판별정확도

1. 서론

1997년 말 외환위기를 전후로 금융기관들의 부실과 제조기업의 부도가 크게 증가하였다. 기업의 파산으로 인한 사회경제적 비용이 매우 크다는 점에서 도산기업과 건전기업을 재무적 특성을 구별하고 이러한 특성을 이용하여 도산을 사전에 예측하는 모형은 매우 중요한 의미를 지니고 있다.

도산예측의 대표적 모형으로는 이항종속변수 회귀모형인 프로빗(probit) 모형과 로짓(logit) 모형이 있다. 프로빗 및 로짓 분석에서는 오차항의 분

포를 각각 표준정규분포와 로지스틱(logistic) 분포로 가정한 후 최우추정법(Maximum Likelihood Estimation: MLE)을 이용하여 추정치를 얻는다. 그러나 이 때 추정치의 일치성(consistency)은 가정한 오차항의 분포가 진정한 오차항의 분포와 차이가 있는 경우에는 담보되지 않는다. 결과적으로 추정치가 진정한 값과 차이를 보이게 되고, 따라서 그러한 추정치에 근거한 모형의 판별력 및 예측력에 문제가 발생할 수 있다.

프로빗 및 로짓 모형의 이러한 제약을 보완하기 위해 본 연구에서는 기존에 채택되어온 분포, 즉 정규분포 및 로지스틱분포에 비해 훨씬 유연한 분

포를 오차항에 도입함으로써 결과적으로 일치추정량을 얻고자 한다. 본 연구에서 사용할 새로운 분포함수는 Johnson(1949a, 1949b)이 제안한 S_U -정규분포이다. S_U -정규분포는 정규분포의 역쌍곡선사인(inverse hyperbolic sine; \sinh^{-1}) 변환을 통해서 만들어지기 때문에 정규분포의 여러 편리한 특성을 그대로 보유하고 있다. 또한 S_U -정규분포는 기존의 정규분포나 로지스틱분포에 비해 훨씬 광범위한 왜도(skewness)와 초과첨도(excess kurtosis)를 나타낼 수 있다.

본 논문은 이항종속변수 회귀모형에 S_U -정규분포를 새롭게 도입했을 뿐만 아니라 다음의 몇 가지 점에서 기존의 연구와 다른 방법론을 취하고 있다. 즉 기존 연구에서는 도산기업과 생존기업의 비율을 1:1로만 한정하여 추정하였으나, 우리는 표본선택 비율을 다양화하면서, 이에 따라 추정치나 판별력이 어떤 차이를 보이는가를 분석한다. 도산 여부를 결정하는 설명변수의 선택에 있어서도 다양한 설명변수의 조합을 제시한다. 또한 기존 연구에서는 독립변수가 도산확률에 어떤 방향으로 유의하게 영향을 미치는가에만 관심이 있었지만 본 연구에서는 추정계수의 부호뿐만 아니라 설명변수가 도산확률에 미치는 한계효과(marginal effect)를 계산함으로써 모형의 실무적 활용도를 높이고자 한다.

본 연구는 다음과 같이 구성되어 있다. 우선 II장에서는 기존 도산예측 모형들의 특징과 차이점을 살펴보고, 본 연구가 이들 연구에서 차지하는 위치와 차별성을 제시한다. 이어 III장에서는 Johnson이 제시한 S_U -정규분포 함수의 특성을 설명하고, S_U -정규분포 함수에 근거한 이항종속변수 모형과 로그우도함수(log-likelihood function)를 도출한다. 또한 도출된 우도함수를 근거로 몬테카를로 시뮬레이션을 이용하여 로짓 모형과 S_U -정규분포 모

형의 추정 성과를 비교분석한다. IV장에서는 실증 분석에 사용될 자료에 대해서 자세히 설명하고, V장에서는 로짓 모형과 S_U -정규분포 모형을 추정한 다음 그 결과를 비교 분석한다. 마지막으로 VI장에서는 연구의 요약과 결론을 제시한다.

II. 도산예측 모형 선행 연구

통계적인 방법으로 기업의 도산을 예측하는 모형은 다변량 판별분석 기법, 이항 종속변수 회귀모형, 옵션가격이론을 이용한 모형, 해저드 모형(hazard model), 인공신경망 모형 등 오래전부터 다양한 방법론들이 제시되어 왔다. 가장 선구적 연구 중의 하나로서 Altman(1968)은 재무비율을 이용한 다변량 판별분석 기법을 이용하여 기업부실을 예측하였으며, 그 후 이와 관련된 많은 후속연구가 이루어져 왔다. 다변량 분석법은 다수의 재무비율을 동시에 고려하는 방법으로서 관찰대상의 여러 개별적인 특성을 이용, 관찰대상을 미리 정해진 선형적 집단들 중 하나로 분류 또는 예측하는데 사용되는 기법이다. 이 방법은 종속변수가 질적인 특성치로 표시되는 경우에 많이 사용된다. 또한 독립변수들이 정규분포임을 가정하고 이러한 가정이 충족되었을 때에만 그 결과를 신뢰할 수 있다. Altman(1968)에 따르면 예측기간이 길어질수록 예측력이 떨어지지만 적어도 부실 1년 전에는 예측력이 비교적 높게 나타나 기업부실의 예측모형으로 활용될 수 있는 것으로 나타났다.

Ohlson(1980)은 다변량 판별분석 등의 기법을 이용한 기존의 연구와는 달리 선형 확률모형인 로짓 모형을 이용하였다. 로짓 모형에서는 판별점수

가 0과 1사이에 존재하여 도산확률을 나타내기 때문에 해석이 용이하다. 다변량 판별분석에 비해 많은 가정이 완화되었으나, 지금까지 대부분의 연구는 오차항의 분포로 정규분포나 로지스틱분포에 의존하는 한계를 지니고 있다.

한편 Merton(1974)은 주식의 수익구조가 콜옵션의 수익구조와 동일하다는 특징을 이용하여 주식의 이론적 가치를 제시했다. 따라서 어떤 기업의 자본이 주식으로만 구성되어 있다는 가정 하에서 주식시가총액 데이터를 이용하여 자산의 확률분포를 추정하는 것이 가능하고, 또한 이를 통해 자산이 부채보다 작을 확률, 즉 도산확률을 추정하는 것이 가능하다는 것을 보였다. 이를 기업 도산확률 추정에 적용한 최근 연구로서 Hillegeist et al. (2004)는 Merton 모형을 이용하여 도산확률을 추정, 기존의 Altman과 Ohlson의 모형결과와 비교 분석하였다. Merton 모형은 이론적으로뿐만 아니라 실무적으로 광범위하게 이용되고 있으나 기본적으로 주가가 존재하는 상장기업만을 분석할 수 있다는 한계가 있다. 또한 Merton 모형은 도산확률만을 제시한다는 점에서 본 논문에서 초점을 맞추는 부분, 즉 주요 재무변수들의 도산확률에 대한 한계효과를 제시하지는 못한다.

판별모형이나 로짓 모형이 도산여부를 종속변수로 설정한데 비해 해저드 모형은 기업의 생존기간(duration)을 종속변수로 설정한다. 이와 관련된 연구로서 Shumway(1999)는 시간의 개념까지 포함시킨 동적(dynamic) 해저드 모형을 이용하여 도산확률을 추정하였다. 그 결과 기존 연구에서 유의하다고 판단되는 설명변수들이 유의하지 않은 경우들이 발견되었다. 그러나 현실적으로 도산가능성

은 도산 직전(1~2년 전)의 재무제표 자료가 중요한 영향을 미치는 것으로 받아들일 때, 시간에 따른 설명변수의 장기적인 변화를 고려한 Shumway 모형의 유용성에 의문이 제기될 수 있다.

인공신경망(neural networks) 모형도 도산예측에 이용되고 있다. 예를 들어 Freeman·Skapura (1991)은 도산예측과 같은 분류문제에 있어서 통계적 방법인 귀납적 학습방법에 비하여 인공신경망 모형이 우수한 인식능력이 있다고 주장했다. 또한 Atiya(2001)의 경우 기존 인공신경망 모형의 문제점을 지적하고 예측의 정확도를 향상시킬 수 있는 방법들을 제시했다. 특히 그는 기업의 주가에 내재된 정보와 예측력을 높이 평가하였다.

이상에서 살펴 본 여러 방법론 중에서, 본 연구는 이항종속변수 회귀모형에 의한 도산예측에 초점을 맞춘다. 국내외를 막론하고 이항종속변수 모형에서 대부분의 선행연구들은 프로빗 혹은 로짓 모형에 의존해 왔다.¹⁾ 그러나 이항 종속변수 모형을 최우추정법으로 추정할 경우 추정치의 일치성은 가정된 오차항의 분포가 진정한 오차항의 분포와 차이가 있는 경우에는 담보되지 않는다. 그런데 정규분포나 로지스틱분포는 그 유연성에 있어서 극히 제약이 심한 분포들이다. 이들은 무엇보다 좌우 대칭분포로서 왜도가 0이다. 또한 분포 꼬리부분의 두터움(혹은 중앙부분의 뽕쪽함)을 나타내는 첨도가 정규분포는 3.0, 로지스틱분포는 4.2로서 고정되어 있다. 두 분포 모두 오로지 단 하나의 왜도-첨도 조합만을 표현할 수 있다는 점에서 유연성에 뚜렷한 한계가 있는 것이다.

프로빗과 로짓 모형의 이러한 제약을 보완하기 위해 본 연구에서는 이들 분포에 비해 훨씬 유연한

1) 이와 관련된 국내연구들로는 신동령(1999, 2005), 이계원(1993), 남주하 외(1995) 등을 참조하라.

S_U -정규분포를 오차항으로 가정함으로써 결과적으로 일치추정량을 얻고자 한다. S_U -정규분포는 기본적으로 좌우비대칭을 나타낼 수 있는 분포로서, 모든 왜도에 대해 로그-정규분포보다 꼬리가 두터운 모든 첨도를 표현할 수 있는 분포이다. 지금까지 계량경제학에 도입된 다양한 비대칭-초과첨도 분포 중에서 가장 광범위한 영역의 왜도-첨도의 조합을 나타낼 수 있는 분포 중 하나이다. 이처럼 유연한 분포를 이용하여 이항종속변수 모형의 일치성 문제를 해결하고자 하는데 본 논문의 의의가 있다. 다음 장에서는 S_U -정규분포가 어떻게 도출되는가를 간단히 살펴보고 이를 이용한 이항종속변수 모형을 제시한 다음 몬테카를로 시뮬레이션을 통해 S_U -정규분포 모형의 상대적 우월성을 점검하기로 한다.

III. S_U -정규분포를 이용한 이항 종속변수 모형

3.1 S_U -정규분포 함수

정규분포는 계량경제학의 이론적·실증적 분석에서 오차항에 대한 가정으로서 중요한 역할을 수행해오고 있다. 그러나 이러한 정규성 가정은 현실적으로 만나게 되는 자료들의 분포를 적절히 표현하지 못하는 경우가 많다. 이러한 한계 때문에 기존의 정규분포 가정을 대체하기 위하여 정규분포에 비해 유연한 분포함수들이 제시되어 왔다(예컨대 McDonald, 1991; McDonald, 1996 등 참조).

Johnson(1949a, 1949b)은 정규분포의 단조변환(monotonic transformation)을 통해 비정규성을 보다 잘 표현할 수 있는 유연한 분포함수를 고안하였다. 정규분포의 단조변환 방식을 이용함으로써 확률분포함수(probability density function: pdf)나 누적분포함수(cumulative distribution function: cdf)의 값을 계산하는 것이 쉬울 뿐만 아니라, 단일변량(univariate) 분포함수를 다변량(multivariate) 차원으로 확장시키는 것도 매우 용이하다는 장점을 지니고 있다. 본 논문에서 우리는 Johnson이 제안하고, 최필선(2005), Choi·Min(2006a) 등에 의해 계량경제학 분석에 도입된 정규분포의 변환방법을 소개하고자 한다.

다음의 조건이 성립하면 확률변수 y 는 S_U -정규분포를 따른다.

$$(1) \sinh^{-1}(y) = \lambda + \theta z$$

여기에서 $-\infty < \lambda < \infty$ 및 $\theta > 0$ 는 각각 왜도와 초과첨도를 표현하는 모수이며, z 는 표준정규확률변수이다. 즉 S_U -정규확률변수에 역쌍곡선사인²⁾(inverse hyperbolic sine)을 취하면 정규확률변수가 된다. 바꾸어 말하면 정규확률변수에 쌍곡선사인을 취하면 S_U -정규확률변수가 된다.

S_U -정규분포의 pdf는 표준정규확률변수의 pdf를 이용하여 다음과 같이 간단히 도출할 수 있다.

$$(2) f(y) = \frac{1}{\theta \sqrt{y^2 + 1}} \phi(y^*)$$

여기에서 $y^* = (\sinh^{-1}(y) - \lambda) / \theta$ 이고 $\phi(\cdot)$ 는 표준정규확률분포의 pdf이다. 또한 S_U -정규분

2) $\sinh^{-1}(y) = \ln(y + \sqrt{y^2 + 1})$ 이며, $\sinh(y) = \frac{e^y - e^{-y}}{2}$ 이다.

포 확률변수의 cdf는 정규분포의 변환이라는 사실 때문에 다음과 같이 간단히 표현된다.

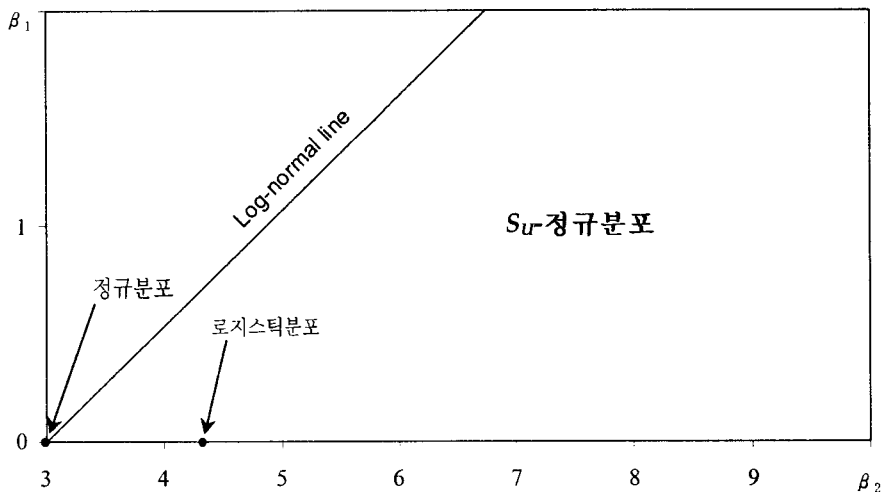
$$(3) F(y) = \Phi(y^*)$$

여기에서 $\Phi(\cdot)$ 는 표준정규확률분포의 cdf이다. S_U -정규분포에서 λ 는 왜도를 결정짓는 모수로서 $\lambda=0$ 일 경우에는 대칭분포(즉, 왜도=0)가 된다. 또한 $\lambda>0$ 이면 오른쪽 꼬리분포이고, $\lambda<0$ 이면 왼쪽 꼬리분포의 모양을 가진다. θ 는 첨도를 결정하는 것으로서 그 값이 클수록 분포의 꼬리가 두꺼워진다. 반대로 $\theta \rightarrow 0$ 이면 S_U -정규분포는 정규분포에 근사하게 된다. 따라서 정규분포는 S_U -정규분포의 극한적 분포(limiting distribution)이다.

<그림 3-1>에서 우리는 S_U -정규분포의 유연성을 확인할 수 있다.(Johnson and Kotz, 1970 참조) 이 그림에서 가로축(β_2)은 첨도이고 세로축(β_1)은 왜도에 제곱을 취한 값을 나타낸다. 우리

는 이 평면상에 어떤 분포함수가 표현할 수 있는 (β_1, β_2) 값을 표시할 수 있는데, 예를 들어 프로빗 모형이 채택하고 있는 정규분포의 경우에는 이 평면의 원점에 해당하는 $(0, 3)$ 이라는 단 하나의 점만 표현 가능하다. 또한 로짓모형이 채택하고 있는 로지스틱분포 역시 $(0, 4.2)$ 의 점만 나타낼 수 있다. 이런 이유 때문에 프로빗 모형과 로짓 모형의 유연성이 극히 낮다고 주장하는 것이다. 한편 우리가 제안하는 S_U -정규분포는 이 그림에서 왼쪽 으로는 로그정규분포선(Log-normal line), 그리고 아래쪽으로는 수평축을 경계선으로 하여 그 이상의 모든 왜도 및 첨도의 조합을 표현할 수 있는 매우 유연한 분포이다. 다시 말하면, 어떤 주어진 왜도에서 로그정규분포보다 꼬리가 두꺼운 모든 분포의 첨도를 S_U -정규분포가 표현할 수 있다는 것이다.

한편, S_U -정규분포에서 왜도와 첨도 모수 외에 위치(location) 모수와 규모(scale) 모수를 추가하



주: 세로축(β_1)은 왜도의 제곱을, 가로축(β_2)은 첨도를 나타냄

<그림 3-1> 왜도·첨도 평면에서 S_U -정규분포의 상대적 유연성

여 일반적인 형태를 만들 수 있는데, 평균이 μ 이고 분산이 σ^2 인 경우— $S_{U}N(\mu, \sigma^2; \lambda, \theta)$ 로 표기—는 다음과 같다.

$$f(x) = \frac{s}{\sigma} f_y\left(\frac{s}{\sigma}(x - \mu) + m\right)$$

여기서 우변의 $f_y(\cdot)$ 는 식 (2)에서 도출된 확률변수 y 의 pdf이고, m 과 s^2 는 S_U -정규분포의 1차 중심적률과 2차 중심적률을 의미한다.³⁾

3.2 S_U -정규분포를 이용한 이항종속변수 모형의 추정

S_U -정규분포를 이항종속변수 모형에 적용할 경우 다음과 같이 표현된다.

$$y_i^* = x_i \beta + \varepsilon_i, \quad \varepsilon_i \sim S_{U}N(0, 1; \lambda, \theta)$$

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases}$$

식 (1)의 변환관계를 이용하면 $\varepsilon_i = (\sinh(\lambda + \theta z_i) - m)/s$ 임을 알 수 있다. 로그우도함수 도출을 위해 다음과 같이 확률 $\Pr(y_i = 0)$ 을 계산할 수 있다.

$$\begin{aligned} (4) \Pr(y_i = 0) &= \Pr(\varepsilon_i \leq -x_i \beta) \\ &= \Pr(z_i \leq (\sinh^{-1}(m - sx_i \beta) - \lambda)/\theta) \\ &= \Phi(a_i^*) \end{aligned}$$

여기에서 $a_i^* = (\sinh^{-1}(m - sx_i \beta) - \lambda)/\theta$ 로 정

의된다. 위 결과를 이용하면 로그우도함수는 다음과 같이 도출된다.

$$\begin{aligned} (5) \log L &= \sum_{y_i=0} \log(\Phi(a_i^*)) \\ &\quad + \sum_{y_i=1} \log(1 - \Phi(a_i^*)) \end{aligned}$$

한편, 본 논문에서는 추정된 계수의 부호뿐만 아니라 설명변수의 단위 변화에 따른 도산확률의 변화, 즉 한계효과에 관심을 갖고 있다. S_U -정규분포 모형에서 각 설명변수의 한계효과는 다음과 같다.⁴⁾

$$\begin{aligned} (6) \frac{\partial \Pr(y_i = 1)}{\partial x_k} &= \frac{s\beta_k}{\theta} \frac{1}{\sqrt{1 + (m - sx_i \beta)}} \phi(a_i^*) \end{aligned}$$

본 연구에서는 식 (5)의 로그우도함수를 극대화하는 (β, λ, θ) 를 GAUSS의 CML-library를 이용하여 추정하였다.

3.3 몬테카를로 시뮬레이션

실제 데이터를 이용한 추정에 앞서 이번 절에서는 몬테카를로(Monte Carlo) 시뮬레이션을 통해 S_U -정규분포 모형과 기존의 로짓 모형의 상대적 우월성을 비교하기로 한다. 이와 비슷한 작업을 Choi · Min(2006b)이 행한 바 있는데, 그들이 프로빗 모형과 S_U -정규분포 모형을 비교한데 반해 여기에서는 학술적 · 실무적으로 더욱 광범위하게 이용되고 있는 로짓 모형을 비교대상으로 한다는

3) y 의 1차~4차 중심적률(central moments)과 그 도출과정은 Choi · Min(2006b)을 참조하라.

4) 본 논문에서 비교대상으로 이용할 로짓 모형의 경우, 특정 설명변수의 한계효과는 $\frac{\partial \Pr(y_i = 1)}{\partial x_k} = \Lambda(x, \beta)[1 - \Lambda(x, \beta)]\beta_k$ 로서 $\Lambda(\cdot)$ 는 로지스틱분포함수의 cdf이다.

차이가 있다. 시뮬레이션 방법은 선행연구들이 수행한 표준적인 방법을 따르기로 한다.

시뮬레이션에 이용할 이항종속변수 모형은 다음과 같은 간단한 구조이다.

$$(7) y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n$$

여기에서 우리는 계수의 참값을 $\beta_0 = -5$, $\beta_1 = 1$ 로 두었으며, 외생변수인 x_i 는 범위가 [0, 10]인 균등분포(uniform distribution)를 따르는 것으로 했다. 또한 오차항인 ε_i 에 대해서는 세 가지 경우를 고려했는데, 정규분포, t 분포, S_U 정규

분포가 그것으로서 모두 평균이 0이고 표준편차가 1인 표준형태를 이용했다. 이 중 t 분포는 꼬리의 두터움을 결정하는 자유도(degrees of freedom) 모수(ν)를 4로 했으며, S_U 정규분포는 λ 와 θ 를 모두 1로 했다. 따라서 t 분포의 경우 대칭분포로서 첨도가 상당히 높은 경우이며, S_U 정규분포는 첨도뿐만 아니라 왜도(skewed to the right)도 비교적 큰 경우이다. 이러한 가정 하에서 샘플의 크기(n)를 1000으로 하여 각 샘플을 로짓 모형과 S_U 정규분포 모형으로 추정했다.⁵⁾

<표 3-1>에는 세 가지 오차항 케이스에 대해 로짓 모형과 S_U 정규분포 모형을 이용하여 추정한

<표 3-1> 시뮬레이션을 통한 추정 성과 비교

	참값	추정값	
		logit 모형	S_U 모형
케이스 1: $\varepsilon_i \sim N(0, 1)$			
$ME(x=3)$	0.054	0.045 (0.013)	0.051 (0.012)
$ME(x=5)$	0.399	0.456 (0.065)	0.407 (0.027)
$ME(x=7)$	0.054	0.045 (.0012)	0.050 (0.012)
$\log L$		-178.48	-176.22
케이스 2: $\varepsilon_i \sim t(0, 1; \nu=4)$			
$ME(x=3)$	0.034	0.044 (0.012)	0.041 (0.016)
$ME(x=5)$	0.530	0.462 (0.147)	0.497 (0.099)
$ME(x=7)$	0.034	0.043 (0.011)	0.038 (0.015)
$\log L$		-179.28	-172.95
케이스 3: $\varepsilon_i \sim S_U N(0, 1; \lambda=1, \theta=1)$			
$ME(x=3)$	0.030	0.033 (0.012)	0.030 (0.006)
$ME(x=5)$	0.521	0.512 (0.084)	0.510 (0.086)
$ME(x=7)$	0.002	0.033 (0.032)	0.002 (0.004)
$\log L$		-159.09	-141.48

주: $ME(x=a)$ 는 $x=a$ 에서의 한계효과를, 그리고 괄호 안은 한계효과 추정치의 평균제곱오차근을 나타낸다.

5) 샘플의 크기나 식 (7)의 계수에 대한 가정들을 달리해도 결과의 기본 내용은 크게 달라지지 않는 것으로 나타났다.

결과가 나와 있다. 각 케이스에 대해 로짓 모형과 S_U -정규분포 모형별로 한계효과에 대한 평균 추정값, 평균제곱오차근(RMSE: root mean squared error), 로그우도값(log L) 등이 나와 있다. 전체적으로 볼 때 S_U -정규분포 모형이 로짓 모형에 비해 추정성 결과가 우월하다는 것을 알 수 있다. 가장 먼저 로그우도값을 보면 세 가지 케이스 모두 S_U -정규분포 모형의 우도값이 더 큰 것을 알 수 있으며, 그 차이는 진정한 오차항 분포가 초과첨도이고 비대칭일수록 더욱 커지는 것으로 나타났다. 또한 한계효과에 있어서도 세 가지 케이스 모두 S_U -정규분포 모형이 로짓 모형에 비해 일부 경우를 제외하고는 추정값이 참값에 더욱 근접할 뿐만 아니라 평균제곱오차근 역시 작다는 것을 알 수 있다.⁶⁾ 특히 진정한 오차항의 분포가 오른쪽 꼬리가 긴 형태인 '케이스 3'의 결과를 보면 로짓 모형의 경우 $x=5$ 에서의 한계효과 추정은 좋은 편이나 $x=7$ 에서의 추정 결과는 크게 악화된 것을 알 수 있다. 이러한 결과는 기본적으로 로짓 모형이 채택하고 있는 로지스틱 분포함수가 대칭분포로서 오차항의 비대칭성을 포착할 수 없다는 점에 기인한다.

IV. 데이터

본 연구의 실증분석을 위한 자료는 증권거래소의 전자공시시스템과 상장기업 공시자료, 상장기업 총람, KIS-VALUE의 재무제표 자료 등을 이용하여

대상 기업을 선정하였다. 분석은 2000년부터 2005년까지 한국증권거래소에 상장된 기업을 대상으로 했다. 이 기업들 중 해당 기간에 상장폐지된 총 60개의 기업을 추출하여 도산기업으로 설정하였다. 상장폐지의 사유는 최종부도, 자본전액 잠식, 영업활동 6개월 이상 정지, 은행거래정지, 회사정리절차 개시 등이다. 한편 이들 도산기업과 함께 추정에 이용될 생존기업은 도산기업과 생존기업의 비율이 1:2 및 1:3이 되도록 무작위로 추출하였다.⁷⁾ 기존의 연구에서는 표본기업의 업종을 제조업이나 특정산업으로 한정하였으나, 본 연구에서는 상장기업의 대부분을 차지하는 제조업, 도소매업, 서비스업을 표본에 포함시켰다. 현실적으로 업종마다 적합한 예측모형을 만든다는 것이 어렵기 때문에 산업더미변수를 설명변수에 포함시켜 업종별 특성을 컨트롤하여 전 산업에 적용 가능한 예측모형을 만들고자 하였다.

도산예측 모형을 위한 설명변수는 기존의 연구를 참고로 하여 도산에 영향을 미칠 것으로 예상되는 재무비율 지표로서 자료의 입수가 가능하고 객관적인 통계처리가 가능한 변수로 선정하였다. 다양한 재무비율 중 본 연구에서는 다음과 같은 선정기준에 의하여 설명변수를 선정하였다.

- (i) 기존 연구문헌들을 검토한 결과 도산기업과 생존기업의 차이를 잘 나타내는 비율로서 기업 부실 연구에서 중요시 되어온 비율
- (ii) 해석이 용이하고 일반적으로 이용되는 비율
- (iii) 재무지표의 값이 극단적으로 크거나 작은

6) 한계효과는 x 의 평균(즉 $x=5$)에서 뿐만 아니라 $x=3$ 및 $x=7$ 등 모두 세 가지 점에서 측정했다. 이는 로짓 모형과 S_U -정규분포 모형이 분포의 평균뿐만 아니라 분포의 전체적인 모양을 제대로 추정하는가를 살펴보기 위한 것이다.

7) 본 연구에서 대응기업을 통한 표본추출보다는 무작위 표본추출 방법을 선택하였으며, 회귀모형에서 총자산과 산업더미 변수를 설명변수에 포함시킴으로써 기업의 규모와 업종 효과를 컨트롤하였다.

유연한 분포함수를 이용한 기업도산예측 모형

값때문에 분석결과를 왜곡 시킬 가능성이 있는 값은 제외

- (iv) 재무지표의 분모는 가급적 총자산을 사용하여 그 값이 비정상적이거나 극단적으로 나타나는 것을 배제⁸⁾

위와 같은 기준을 적용하여 <표 4-1>의 재무비율과 산업 더미변수를 모형의 설명변수로 선정하였

다. 다중공선성(multi-collinearity)문제를 고려하여 (운전자본/총자산)과 (유동부채/유동자산) 변수 중 하나, 그리고 (EBIT/총자산)과 ROA 변수 중 하나 만을 모형에 포함시켰다.⁹⁾

현금지급능력 변수는 전통적으로 도산예측 모형에서 가장 유의한 변수 중의 하나로 다루어져 왔으며, 실무적으로 매우 중요시하는 재무비율이다. 이는 이익변수와 달리 회계처리 방법의 영향을 덜 받

<표 4-1> 설명변수 정의

재무지표의 성격	변수	비고
산업	제조업 (x11) 도매 및 소매업 (x12) 사업서비스업 (x13)	- 가변수(dummy variable) 형태 - x12와 x13만 모형에 포함
자본구조 관련비율	총부채/총자산 (x2)	
단기지급능력	운전자본/총자산 (x31) 유동부채/유동자산 (x32)	둘 중 한 변수 사용
현금지급능력	영업활동 현금흐름/총부채 (x4)	
수익성비율	이자 및 세금 차감 전 이익(EBIT)/총자산 (x51) 당기순이익/총자산: ROA (x52) 이익잉여금/총자산 (x53)	둘 중 한 변수 사용
기업규모	log(총자산) (x6)	

<표 4-2> 도산기업 데이터의 연도별 도산사유

	2000	2001	2002	2003	2004	2005	계
부도발생 또는 은행거래정지	4	0	7	6	8	5	30
회사정리절차 개시신청	0	0	0	0	1	1	2
자본전액잠식	6	3	6	3	3	6	27
영업활동정지 6개월 이상	0	1	0	0	0	0	1
계	10	4	13	9	12	12	60

8) 예를 들어 자본잠식된 기업의 재무구조를 (총부채/총자본)으로 측정하면 (-)값으로 나타나 분석결과를 왜곡시키지만, (총부채/총자산)으로 측정하면 이와 같은 문제가 발생하지 않는다.

9) 다중공선성 문제를 확인하기 위해 변수간 상관계수를 점검한 결과, 운전자본/총자산(x31)과 유동부채/유동자산(x32)의 상관계수는 -0.79이고, EBIT/총자산(x51)과 ROA(x52)변수간의 상관계수는 0.98로서 모두 0과 유의하게 다른 것으로 나타났다. 또한 로짓 모형에서 x51과 x52변수를 모두 포함시켰을 때, x51변수의 표준오차는 12.22로 x51변수만 포함시켰을 때의 표준오차 1.16보다 12배 정도 큰 것으로 나타났다.

고, 분식이 용이하지 않을 뿐만 아니라, 부채상환 능력과 배당지급능력을 파악하는데 유용한 변수로 나타나기 때문이다. 본 연구에서는 영업활동 현금흐름을 이용하여 기업의 동태적 현금상황 변동을 측정한다. 당기순이익을 이용한 수익성 관련지표는 전통적 성과지표로서 기업평가·분석에서 가장 널

리 사용되는 변수이다. 또한 누적 수익성을 대표하는 변수로는 (이익잉여금/총자산)을 포함시켰다.

우선 표본으로 선정된 도산기업의 도산 사유별 분포는 <표 4-2>와 같다. 여기에서 보듯이 부도발생 또는 은행거래 정지의 사유로 도산한 기업이 전체의 50%로 가장 많고, 그 다음이 자본전액잠식

<표 4-3> 설명변수 데이터의 기초통계량

변수명	기업구분	평균	분산	최대값	중앙값	최소값	t값
총부채 /총자산	도산기업	1.55	2.29	11.65	1.16	0.35	
	생존기업(1:2)	0.48	0.07	1.90	0.45	0.06	-7.54
	생존기업(1:3)	0.48	0.07	1.90	0.45	0.06	-9.08
운전자본 /총자산	도산기업	-0.21	0.25	0.55	-0.13	-2.02	
	생존기업(1:2)	0.23	0.06	0.89	0.21	-0.94	8.00
	생존기업(1:3)	0.21	0.07	0.89	0.20	-0.94	8.33
유동부채 /유동자산	도산기업	1.69	2.03	8.17	1.29	0.14	
	생존기업(1:2)	0.67	0.38	4.69	0.58	0.05	-6.73
	생존기업(1:3)	0.72	0.39	4.69	0.61	0.05	-7.29
영업활동 현금흐름 /총부채	도산기업	-0.02	0.03	0.30	-0.01	-0.79	
	생존기업(1:2)	0.17	0.24	2.57	0.13	-1.82	3.12
	생존기업(1:3)	0.19	0.24	2.57	0.13	-1.82	3.37
EBIT /총자산	도산기업	-0.24	0.28	1.10	-0.16	-2.25	
	생존기업(1:2)	0.11	0.09	2.21	0.07	-1.17	5.64
	생존기업(1:3)	0.09	0.09	2.21	0.07	-1.17	6.03
당기순이익 /총자산: ROA	도산기업	-0.30	0.26	1.02	-0.21	-2.32	
	생존기업(1:2)	0.07	0.07	1.64	0.05	-1.33	6.43
	생존기업(1:3)	0.06	0.08	1.64	0.04	-1.33	6.92
이익잉여금 /총자산	도산기업	-0.68	0.66	0.18	-0.45	-4.11	
	생존기업(1:2)	0.01	0.25	1.39	0.02	-3.13	7.05
	생존기업(1:3)	0.02	0.20	1.52	0.02	-3.13	8.47
log(총자산)	도산기업	10.78	1.07	13.37	10.56	7.95	
	생존기업(1:2)	10.83	2.00	16.68	10.61	8.64	0.25
	생존기업(1:3)	10.70	1.87	16.68	10.53	8.50	-0.12

순으로 나타났으며, 영업활동 정지 6개월 이상이 가장 낮은 빈도를 보였다.

앞서 언급하였듯이 이항종속변수 모형에서 오차항의 가정은 추정치의 일치성에 영향을 미친다. 이와 관련하여 Choi·Min(2006b)은 이항종속변수 모형에서 오차항의 분포가 정규분포를 따르는가를 검정하는 LR(likelihood ratio) 검정을 제안하였다. 우리의 데이터에 대해 그들이 제안한 방법을 응용하여, 귀무가설을 로지스틱 분포로 설정한 후, 검정한 결과 위의 모든 경우에 오차항이 로지스틱 분포라는 귀무가설을 기각하는 것으로 나타났다. 이는 로짓 모형의 추정치가 일치추정량이 되지 않을 수 있다는 것을 의미한다.

〈표 4-3〉은 설명변수에 대한 기초통계량, 도산기업과 생존기업 간의 모평균차이를 검정하는 t -검정 결과를 제시하고 있다. 도산기업은 도산 1년 전의 재무비율 데이터를 이용하였으며¹⁰⁾, 생존기업의 표본은 각 년도의 도산기업의 수에 비례하여 1:2, 1:3의 두 가지 비율로 무작위 추출하였다. 각 변수의 t -검정 결과를 살펴보면 기업 규모를 나타내는 $\log(\text{총자산})$ 을 제외하고는 모든 설명변수에 대해 도산기업과 생존기업의 평균값이 유의한 차이를 보여준다. t 값의 부호를 보면 (총부채/총자산), (유동부채/유동자산)의 경우에는 음의 값을 보여 예상 부호와 일치하며, 그 밖의 모든 설명변수에서는 t 값이 양으로서 이 역시 예상 부호와 일치하는 것으로 여겨진다.

V. 실증분석 결과

5.1 추정결과 비교분석

여기서는 이상의 데이터를 이용하여 기업의 도산 예측 모형으로서 S_U -정규분포 모형의 추정결과 및 판별력을 로짓 모형과 비교·분석한다. 또한 앞서 언급한대로 설명변수의 조합과 표본선택비율을 변화시키면서 추정결과가 어떻게 달라지는가를 분석하기로 한다.

〈표 5-1〉과 〈표 5-2〉는 표본선택비율에 따른 로짓과 S_U -정규분포(S_{UN}) 모형의 계수 추정 결과를 비교해서 보여주고 있다. 〈표 5-1〉에서는 현금지급능력을 나타내는 변수로서 (운전자본/총자산), 그리고 수익성 변수로서 ROA 변수를 모형에 포함시켰다.¹¹⁾ 우선 계수의 유의성을 살펴보면, 표본선택비율에 상관없이 두 모형의 유의성 결과에 거의 차이가 없음을 알 수 있다. 즉 (총부채/총자산)과 ROA 변수 등이 유의하며, 추정된 부호는 예상부호와 모두 일치한다. 특히 ROA 변수가 5% 수준에서 모두 유의하며, 예상부호인 (-)로 추정되었다. 주목할 만한 것은 (이익잉여금/총자산) 변수의 경우 로짓 모형에서는 (+)로 유의한 것으로 나타나 예측에 반대되는 결과를 제시하는 반면, S_U -정규분포의 경우에는 동 변수의 유의수준이 떨어질 뿐만 아니라 표본선택비율이 1:3이 되면 유의하지 않은 것으로 나타났다. 규모변수인 $\log(\text{총자산})$ 은 오차항 가정과 표본선택비율에 상관없이 항상 유의

10) 본 연구에서는 도산 2년 전 자료도 표본 추출하여 모형을 추정하였다. 그러나 도산여부에 대한 판별력이 낮아 2년 전 자료의 기초통계량은 제시하지 않는다.

11) 추정모형 1의 변형모형으로 ROA대신 (EBIT/총자산)을 포함한 모형을 추정하였으나, 모형 1의 결과와 큰 차이를 보이지 않아 추정결과를 제시하지 않는다.

〈표 5-1〉 로짓 모형과 S_U -정규분포 모형 추정결과: 모형 1

	도산기업 대 생존기업 비율			
	1:2		1:3	
	logit	$S_U N$	logit	$S_U N$
상수항 (constant)	-3.453 (0.159)	-1.159 (0.732)	-6.344** (0.003)	-1.321** (0.574)
도매 및 소매업 (x12)	-0.219 (0.822)	-0.288 (0.258)	-0.346 (0.674)	-0.204 (0.183)
사업서비스업 (x13)	-1.264 (0.495)	-0.266 (0.716)	-0.835 (0.570)	-0.208 (0.419)
총부채/총자산 (x2)	5.064** (0.000)	1.347** (0.415)	4.869** (0.000)	0.915** (0.361)
운전자본 /총자산 (x31)	-0.941 (0.384)	-0.519 (0.342)	0.021 (0.981)	-0.076 (0.248)
영업활동 현금흐름 /총부채 (x4)	-0.296 (0.818)	-0.338 (0.450)	0.181 (0.855)	-0.047 (0.239)
ROA (x52)	-3.801** (0.001)	-1.070** (0.443)	-3.182** (0.001)	-0.674** (0.367)
이익잉여금 /총자산 (x53)	2.135** (0.006)	0.543* (0.330)	1.447** (0.027)	0.252 (0.228)
log(총자산) (x6)	-0.099 (0.644)	-0.003 (0.063)	0.128 (0.473)	0.029 (0.037)
log L	-51.767	-49.235	-66.074	-62.796

주: **, *은 각각 유의수준 5%와 10%에서 유의함을 의미하며 괄호안은 표준오차임.

하지 않은 것으로 나타났다.

〈표 5-2〉에서는 〈표 5-1〉모형에서 포함된 (운전자본/총자산) 대신 (유동부채/유동자산)변수를 포함시키고, 수익성을 대표하는 변수로서 ROA변수 대신 (EBIT/총자산)을 포함시켰다. 이 경우 〈표 5-1〉과 유사하게 단기지급능력을 나타내는 (유동부채/총자산)은 유의하지 않음을 알 수 있다. 〈표 5-1〉의 결과와 유사하게 (EBIT/총자산)변수 역시 유의하며, 예상부호인 (-)로 추정됨을 알 수 있다.

이상의 결과를 요약하면 로짓 모형과 S_U -정규분포 모형은 추정 계수의 값은 다르지만 유의성에서는 큰 차이가 없다고 할 수 있다.

〈표 5-3〉과 〈표 5-4〉는 각 설명변수의 도산확률에 대한 한계효과를 모형별로 계산한 결과이다. 한계효과란 관심변수 이외의 모든 설명변수의 값을 표본평균으로 고정시킨 상태에서 관심변수의 도산확률에 대한 편미분 값을 의미한다. 우선 〈표 5-3〉의 결과를 살펴보면, 표본비율이 1:2일 때,

〈표 5-2〉 로짓 모형과 S_U -정규분포 모형 추정결과: 모형 2

	도산기업 대 생존기업 비율			
	1:2		1:3	
	logit	S_U	logit	S_U
상수항 (constant)	-4.041* (0.091)	-1.286* (0.770)	-6.424** (0.002)	-0.964** (0.404)
도매 및 소매업 (x12)	-0.345 (0.725)	-0.302 (0.259)	-0.420 (0.612)	0.035 (0.121)
사업서비스업 (x13)	-1.727 (0.434)	-0.518 (0.992)	-1.057 (0.512)	-0.231 (0.456)
총부채/총자산 (x2)	5.541** (0.000)	1.523** (0.449)	5.011** (0.000)	0.874** (0.339)
유동부채 /유동자산 (x32)	0.228 (0.468)	0.132 (0.140)	0.093 (0.697)	-0.016 (0.045)
영업활동 현금흐름 /총부채 (x4)	-0.103 (0.937)	-0.200 (0.439)	0.362 (0.717)	0.070 (0.199)
EBIT/총자산 (x51)	-4.282** (0.000)	-1.231** (0.452)	-3.382** (0.000)	-0.615* (0.340)
이익잉여금 /총자산 (x53)	2.316** (0.004)	0.614* (0.325)	1.566** (0.018)	0.226 (0.217)
log(총자산) (x6)	-0.091 (0.678)	-0.015 (0.070)	0.127 (0.478)	0.003 (0.031)
log L	-50.428	-48.240	-64.901	-61.735

주: **, *은 각각 유의수준 5%와 10%에서 유의함을 의미하며 괄호안은 표준오차임.

로짓 모형에서는 (총부채/총자산)의 비율이 10%p 상승하면 도산확률이 11.15%p 유의하게 상승하는 반면, S_U -정규분포 모형에서는 도산확률이 8.86%p 상승하는 것으로 나타났다. 즉 S_U -정규분포 모형에서 (총부채/총자산)이 도산확률에 미치는 영향이 작다는 것을 의미한다. 관심변수인 ROA를 살펴보면 표본비율이 1:2일 때, ROA비율 10%p 상승시 로짓 모형에서는 도산확률이 8.36%p 하락하고, S_U -정규분포 모형에서는 도산확률이 7.04%p 하

락한다. 한편 (이익잉여금/총자산) 변수의 경우, 표본비율이 1:3일 때 S_U -정규분포 모형에서는 로짓 모형과 달리 한계효과가 유의하지 않다는 것도 주목할 만하다.

단기지급능력의 대리변수로 (유동부채/유동자산)과 수익성변수로 (EBIT/총자산)을 포함시킨 〈표 5-4〉의 결과를 살펴보면, 표본비율이 1:3인 경우, (EBIT/총자산)의 비율이 10%p 상승할 때, 로짓 모형의 경우 도산확률이 5.05%p 하락하고, S_U -정

〈표 5-3〉 설명변수의 한계효과: 모형 1

	도산기업 대 생존기업 비율			
	1:2		1:3	
	logit	<i>SuN</i>	logit	<i>SuN</i>
도매 및 소매업 (x12)	-0.046	-0.119	-0.048	-0.089
사업서비스업 (x13)	-0.211	-0.107	-0.099	-0.085
총부채/총자산 (x2)	1.115**	0.886**	0.747**	0.687**
운전자본 /총자산 (x31)	-0.207	-0.342	0.003	-0.057
영업활동 현금흐름 /총부채 (x4)	-0.065	-0.222	0.027	-0.035
ROA (x52)	-0.836**	-0.704**	-0.488**	-0.506**
이익잉여금 /총자산 (x53)	0.470**	0.357*	0.222**	0.189
log(총자산) (x6)	-0.021	-0.002	0.019	0.022

주: 1) **, *은 각각 유의수준 5%와 10%에서 유의함을 나타냄.

2) x12와 x13 변수에 대한 한계효과는 $\Pr(y_i=1 | x_i=0) - \Pr(y_i=1 | x_i=1)$ 로 계산.

규분포 모형에서는 9.59%p 하락하는 것으로 나타났다. 이상에서 보듯이 한계효과 크기 측면에서는 로짓과 *SuN*-정규분포 모형의 추정결과에 상당한 차이가 있음을 알 수 있다.

〈표 5-5〉에서는 추정된 계수가 추정방법에 따라 유의한 차이를 보이는지 여부를 보수적(conservative) *t*-test를 사용하여 분석했다. 일반적인 *t*-test 대신 보수적 *t*-test를 사용한 이유는 같은 표본을 이용하여 서로 다른 방법으로 추정된 결과를 비교함에 따라 사용된 표본이 서로 독립이라는 가정을 만족하지 않기 때문이다.(Min · Kim, 2003 참조)

〈표 5-5〉에서는 모형 1과 모형 2에서 유의하게 추정된 설명변수를 중심으로 검정결과를 제시하고 있다. 로짓 모형과 *SuN*-정규분포 모형에 의해 추정된 계수는 모두 5% 유의수준에서 서로 유의하게 차이를 알 수 있다. 이러한 차이는 표본선택비율과 추정모형을 달리하는 경우에도 여전히 발견되고 있다. 따라서 두 추정방법에 의해 도출된 계수는 유의하게 다르며, 또한 추정계수를 통해 산출되는 한계효과 역시 추정방법에 따라 유의하게 달라질 것으로 예상할 수 있다.

〈표 5-4〉 설명변수의 한계효과: 모형 2

	도산기업 대 생존기업 비율			
	1:2		1:3	
	logit	<i>SuN</i>	logit	<i>SuN</i>
도매 및 소매업 (x12)	-0.071	-0.129	-0.056	0.063
사업서비스업 (x13)	-0.256	-0.158	-0.114	-0.151
총부채/총자산 (x2)	1.219**	1.052**	0.748**	1.362**
유동부채 /유동자산 (x32)	0.050	0.091	0.013	-0.025
영업활동 현금흐름 /총부채 (x4)	-0.022	-0.138	0.054	0.109
EBIT/총자산 (x51)	-0.942**	-0.850**	-0.505**	-0.959*
이익잉여금 /총자산 (x53)	0.509**	0.424*	0.234**	0.352
log(총자산) (x6)	-0.020	-0.010	0.019	0.005

주: 1) **, *은 각각 유의수준 5%와 10%에서 유의함을 나타냄.

2) x12와 x13 변수에 대한 한계효과는 $\Pr(y_i=1 | x_i=0) - \Pr(y_i=1 | x_i=1)$ 로 계산.

〈표 5-5〉 로짓 모형과 *Su*-정규분포 모형 추정계수에 대한 보수적 t-test

설명변수	모형 1		설명변수	모형2	
	1:2	1:3		1:2	1:3
총부채/총자산 (x2)	8.956**	10.952**	총부채/총자산 (x2)	8.948**	12.203**
ROA (x52)	-6.164**	-6.833**	EBIT/총자산 (x51)	-6.750**	-8.138**
이익잉여금 /총자산 (x53)	4.823**	5.204**	이익잉여금 /총자산 (x53)	5.236**	6.153**

주: 1) **, *은 각각 유의수준 5%와 10%에서 유의함을 나타냄.

2) 보수적 t-test 통계량 $T_c = \frac{\hat{\beta}_L - \hat{\beta}_S}{\sqrt{2[\text{var}(\hat{\beta}_L) + \text{var}(\hat{\beta}_S)]}}$, 여기서 $\hat{\beta}_L$ 와 $\hat{\beta}_S$ 는 각각 로짓 모형과 *Su*-정규분포에서 추정계수를 의미함.

5.2 판별정확도 및 예측정확도 분석

도산예측 모형의 유용성은 추정모형이 도산기업과 생존기업을 얼마나 정확하게 분류하는가를 측정하는 판별정확도(classification rate)와 예측정확도(prediction rate)를 이용하여 판단할 수 있다.¹²⁾ 판별정확도를 계산하기 위해, 앞서 설정한 2개의 설명변수 조합 모형의 추정결과를 이용하여 도산확률을 계산하였다. 제1종 오류와 제2종 오류¹³⁾의 상대적 오류비용을 1:1로 설정하고 예측된 도산확률이 0.5보다 크면 도산기업으로, 0.5보다 작으면 생존기업으로 분류했다. <표 5-6>은 도산기업과 생존기업의 판별정확도 결과를 보여주고 있다.¹⁴⁾ 이 표에서 보듯이 제 1종 오류의 경우, 2개 모형 모두에서 S_U -정규분포 모형의 판별정확도가 로짓 모형보다 우수한 것으로 나타났다. 또한 이와 같은 정확도의 차이는 표본비율이 증가할수록 커짐을 알 수 있다.

현실적으로 볼 때 제1종 오류비용이 제2종 오류

비용보다 훨씬 크기 때문에 <표 5-7>에서는 상대적 오류비용을 고려한 확률임계치(cut-off probability)를 결정하여 예측정확도를 판단하였다. Beneish (1999)에서와 같이 제1종 오류와 제2종 오류의 상대적인 비용을 1:2~1:10까지 가정하고, 추정표본을 이용하여 분류오류에 따른 기대비용을 최소화하는 확률임계치를 찾아냈다. 그리고 주어진 확률임계치에 따른 검증표본에서의 분류오류를 계산하였다. 총기대비용은 다음과 같이 전체 표본에서 도산기업이 차지하는 비율과 상대적 오류비용을 고려하여 계산하였다.

$$\text{총기대비용} = P(M) \cdot P_1 \cdot C_1 + (1 - P(M)) \cdot P_2 \cdot C_2$$

위 식에서 $P(M)$ 은 전체 표본에서 도산기업이 차지하는 비율이고, P_1 과 P_2 는 주어진 확률임계치에서 제1종 및 제2종의 상대적 오류비용을 나타내는 조건부 확률이다. 그리고 C_1 과 C_2 는 제1

<표 5-6> 판별정확도(상대적 오류비용 1:1)

		도산기업 대 생존기업 비율			
		1:2		1:3	
		logit	SUN	logit	SUN
모형 1	제 1종 오류	0.183	0.150	0.283	0.183
	제 2종 오류	0.050	0.050	0.039	0.050
모형 2	제 1종 오류	0.167	0.150	0.267	0.200
	제 2종 오류	0.042	0.042	0.039	0.061

12) 판별정확도는 추정표본만을 이용하여 도산기업으로의 분류정확도를 계산하였고, 예측정확도는 추정표본과 검증표본으로 나누어 추정표본의 추정결과에 기초하여 검증표본의 분류정확도를 계산하였다.
 13) 제 1종 오류는 도산기업을 생존기업으로 분류하는 오류이고, 제 2종 오류는 생존기업을 도산기업으로 분류하는 오류를 뜻한다.
 14) Dopuch · Holthausen · Leftwich(1986)에 따르면, 표본선택비율이 증가할수록 제1종 오류의 값이 증가하고 제2종 오류의 값은 감소한다. 이러한 편의(bias) 문제는 표본선택비율에 따른 추정계수의 편의에 기인하는 것으로 보인다.

〈표 5-7〉 상대적 오류비용에 따른 예측정확도

	모형 1				
		logit		$S_U N$	
오류비용	확률임계치	1종오류	2종오류	1종오류	2종오류
2:1	0.740	0.483	0.016	0.350	0.028
3:1	0.740	0.483	0.016	0.350	0.028
4:1	0.740	0.483	0.016	0.350	0.028
5:1	0.740	0.483	0.016	0.350	0.028
10:1	0.480	0.250	0.027	0.183	0.028
	모형 2				
		logit		$S_U N$	
오류비용	확률임계치	1종오류	2종오류	1종오류	2종오류
2:1	0.740	0.450	0.016	0.383	0.017
3:1	0.740	0.450	0.016	0.383	0.017
4:1	0.740	0.450	0.016	0.383	0.017
5:1	0.740	0.450	0.016	0.383	0.017
10:1	0.460	0.233	0.027	0.183	0.026

주: 표본선택비율은 1:3을 설정함.

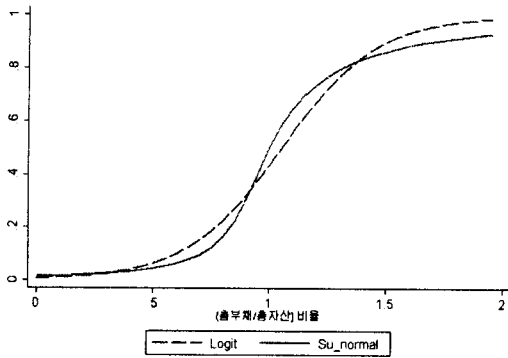
종 오류와 제2종 오류비용을 의미한다. 모든 모형에서 제1종 오류의 경우, S_U -정규분포 모형이 로짓 모형보다 더 우수한 예측정확도를 보여주고 있으며, 그러한 차이는 모형 1에서 더 커짐을 알 수 있다.

5.3 도산확률에 대한 시뮬레이션

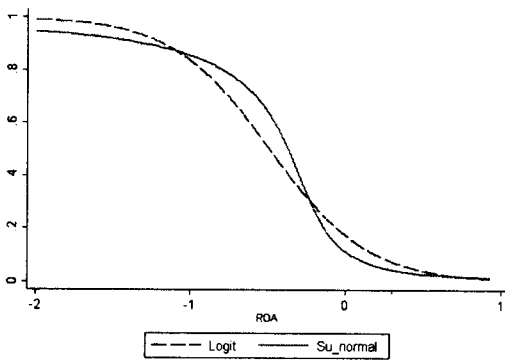
본 절에서는 관심변수 이외의 다른 변수는 표본 평균에서 고정시킨 후 관심변수의 값을 조금씩 변화시킬 때 도산확률이 어떻게 변하는가를 계산하여 이를 그래프로 나타내었다. 〈그림 5-1〉부터 〈그림 5-3〉까지는 〈표 5-1〉에 제시된 모형 1에서 표본 선택비율이 1:3인 경우를 택해 그 결과를 시뮬레

이션으로 나타낸 그림들이다. 먼저 〈그림 5-1〉은 (총부채/총자산)의 변화에 따른 도산확률의 변화를 보여주고 있다. 여기에서 보듯이 (총부채/총자산)의 비율이 0.5~1.0 사이에 있는 경우 로짓 모형의 도산확률이 S_U -정규분포 모형에 비해 더 낮지만, 1.0~1.5 사이에서는 S_U -정규분포 모형의 도산확률이 오히려 더 높다는 것을 알 수 있다. 로짓 모형의 경우 낮은 부채비율을 가진 기업에 대해서 S_U -정규분포 모형에 비해 더 높은 도산확률을 부여하고 있음을 의미한다.

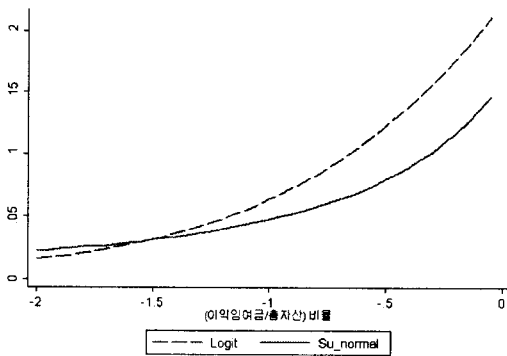
〈그림 5-2〉는 ROA 변수의 변화에 따른 도산확률을 보여주고 있다. 이 비율이 -1.0~-0.2 사이에 있으면 S_U -정규분포 모형의 도산확률이 로짓 모형에 비해 더 높다는 것을 알 수 있다. 마지막으로



〈그림 5-1〉 총부채/총자산 (x2) 변화에 따른 도산확률



〈그림 5-2〉 ROA (x51) 변화에 따른 도산확률



〈그림 5-3〉 이익잉여금/총자산 (x53) 변화에 따른 도산확률

로 〈그림 5-3〉은 (이익잉여금/총자산) 변수의 변화에 따른 도산확률을 그린 그래프이다. 앞서 설명하였듯이 로짓 모형에서 이 변수는 (+)의 한계효과로 인해, 비율이 증가할수록 도산확률이 증가한다. 그러나 SU -정규분포 모형의 경우에는 도산확률이 증가하기는 하지만, 이러한 증가가 유의하지 않음을 알 수 있다.

VI. 맺음말

본 연구에서는 외환위기 이후인 2000~2005년 기간 중 도산한 기업과 생존기업의 재무비율을 이용하여 도산예측 모형을 구축했다. 기존의 이항종속변수 회귀모형에서는 오차항으로 대부분 정규분포나 로지스틱분포와 같이 유연성에 제약이 심한 분포를 사용한다. 반해, 본 연구에서는 유연성이 매우 높아 추정치의 일치성을 담보할 수 있는 SU -정규분포 함수를 도입하였다. 또한 도산기업과 생존기업의 비율을 1:2와 1:3 두 가지로 선택하여 추정결과를 비교·분석하였을 뿐만 아니라, 도산확률에 영향을 미치는 설명변수를 다양하게 조합하여 한계효과에 대한 비교분석을 행하였다.

추정결과 로짓 모형과 SU -정규분포 모형은 추정계수의 유의성 자체에서는 큰 차이를 보이지 않았지만, 설명변수의 한계효과 측면에서는 상당한 차이를 보였다. 우선 두 모형 모두 자본구조 관련비율인 (총부채/총자산) 변수는 유의하게 (+)의 영향을 보이고, 수익성 비율인 (EBIT/총자산)과 ROA 변수는 유의하게 (-)의 영향을 미치는 것으로 추정되었다. 그러나 SU -정규분포 모형에서는 로짓 모형과 달리 표본비율이 커질수록 (이익잉여

금/총자산) 변수에 있어서 (+)의 유의성이 점차 사라지는 것으로 나타났다. 한편 추정계수에 근거하여 도산여부에 대한 판별정확도 및 예측정확도를 계산한 결과, S_U -정규분포 모형이 로짓 모형에 비해 우월한 판별력 및 예측력을 지니는 것으로 나타났다. 이와 같은 결과와 앞에서 언급한 한계효과에서 두 모형간의 큰 차이를 종합적으로 고려하면, 로짓 모형의 경우 제약적인 분포함수 가정으로 인해 추정치의 일치성이 담보되지 않음으로써 한계효과 추정치에 큰 차이를 유발시키는 것은 물론 판별력과 예측력을 저하시키는 것으로 해석할 수 있다.

이처럼 오차항에 대한 기존의 제약적 가정을 완화한 새로운 최우추정법을 통하여 관심 있는 재무비율이 도산 가능성에 미치는 영향과 도산확률에 미치는 한계효과를 보다 정확하게 추정함으로써 실무적 활용도를 높일 수 있을 것으로 생각된다.

참고문헌

- 남주하 외(1995), "부도예측모형 분석," 한국경제연구원 연구보고서, 1-89.
- 신동령(1999) "도산기업의 재무적 특징과 도산예측모형" **한일경상논집**, 17권, 131-156.
- 신동령(2005), "부실기업의 재무적 특징과 부실예측모형에 관한 연구," **회계정보연구**, 23권 2호, 137-165.
- 이계원(1993), "회계정보에 의한 기업도산예측과 시장반응," **회계학연구**, 16호, 49-77.
- 최필선(2005), "새로운 다변량 분포함수를 이용한 GARCH 모형 추정," **재무연구**, 18권 2호, 185-208.
- Altman, E.(1968) Financial Ratios, Discriminant Analysis and Prediction of Corporate Bankruptcy." *Journal of Finance*, Vol.23 No.4, 589-609.
- Atiya, A. F. (2001), "Bankruptcy Prediction for Credit Risk Using Neutral Networks: A Survey and New Results," *IEEE Transactions on Neural Networks*, 12(4), 929-935.
- Beneish, M. D. (1999), "The Detection of Earnings Manipulation" *Financial Analysis Journal* 55(September/October): 24-36.
- Choi, P. and I. Min(2006a), "Estimating Endogenous Switching Regression Model with a Flexible Parametric Distribution Function: Application to Korean Housing Demand," Forthcoming in *Applied Economics*.
- _____ (2006b), "Likelihood Inference and Conditional Normality Test with Johnson's S_U -normal Distribution," Working Paper, Kyung Hee University.
- Dopuch N., R. Holthausen, and R. Leftwich (1986), "Abnormal Stock Returns Associated with Media Disclosures of 'Subject to' Qualified Audit Opinions," *Journal of Accounting and Economics*, Vol.8 Iss.2, 93-117.
- Freeman, J. A. and D. M. Skapura(1991), "Algorithms, Applications and Programming Techniques," in *Neural Network*, Addison-Wesley Publishing Company.
- Hillegeist, S.A., Keating, E.K., Cram, P., and Lundstedt, K.G.(2004), "Assessing the Probability of Bankruptcy," *Review of Accounting Studies*, Vol. 9 No.1: 5-34
- Johnson, N. L.(1949a), "Systems of Frequency Curves Generated by Method of Translation," *Biometrika*, 36, 149-176.
- _____ (1949b), "Bivariate Distributions Based on Simple Translation Systems," *Biometrika*,

36. 297-304.
- Johnson, N. L., and S. Kotz(1970), *Continuous Univariate Distributions*, J. Wiley, New York.
- McDonald, J. B.(1991), "Parametric Models for Partially Adaptive Estimation with Skewed and Leptokurtic Residuals," *Economics Letters*, 37, 273-278.
- _____ (1996), "An Application and Comparison of Some Flexible Parametric and Semi parametric Qualitative Response Models," *Economics Letters*, 53, 145-152.
- Merton, R.(1974), "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," *Journal of Finance*, Vol. 29 No. 2, 449~470.
- Min, I. and Kim, J. (2003), "Modeling Credit Card Borrowing: A Comparison of Type I and Type II Tobit Approaches". *Southern Economic Journal*: Vol. 70, No. 1, pp. 128-143.
- Ohlson, J.(1980), "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, Vol.18 No.1, 109-131.
- Shumway, T. (1999) "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *The Journal of Business* Vol.74 No.1: 101-124.

Bankruptcy Prediction Model with a Flexible Parametric Distribution Function

Pilsun Choi* · Mihyun Cha** · Insik Min***

Abstract

Research on bankruptcies from business failures is very important issue academically as well as practically. Even though many empirical studies have tried to propose certain kinds of predictive models for bankruptcy, they show various estimation results depending on the econometric methods, the explanatory variables, and the sample selection criteria. In this paper, we propose a new and more flexible S_U -normal distribution function as the error distribution in the binary dependent variable model. The S_U -normal distribution introduced in this study can provide the consistent coefficient estimates and thus yield more accurate bankruptcy predictions. Using Monte Carlo simulations, we show that S_U -normal distribution performs better than logistic distribution in terms of capturing the non-normality of error term. During the period 2000~2005, estimation comparisons using listed companies show that the performance of S_U -normal model is better than that of logit model in terms of classification rate as well as prediction rate. In addition, the empirical results demonstrate that there is substantial difference between two models in the marginal effects. The main contribution of this paper is that the S_U -normal model can provide more precise quantification for the bankruptcy probability and the marginal effects of the explanatory variables observed in financial statements.

Key words: S_U -normal distribution, Bankruptcy prediction model, Classification rate

* Assistant Professor, Department of International Trade Konkuk University

** Ph.D Candidate, Department of Accounting Kyung Hee University

*** Assistant Professor, Department of Economics Kyung Hee University