

기대효용최대화를 통한 한국형 기업 신용평가 모형*

박유성¹⁾ 송지현²⁾ 최보승³⁾

요약

기업의 신용을 평가하는데 있어 정확한 파산확률의 추정은 무엇보다도 중요한 요소이다. 선형로지스틱회귀모형보다 성능이 좋은 기대효용최대화 (Maximum Expected Utility) 모형이 제안되었다. 그러나 이 모형에 포함되어 있는 모수의 일부가 북미와 유럽지역의 자료를 토대로 경험적으로 추정되어진 것이므로 우리나라 기업에 바로 적용하기에는 무리가 있다. 따라서 우리나라 중소기업의 자료를 바탕으로 모수를 재추정하여 한국형 MEU모형을 제안하고자 한다. 34,057개의 중소기업을 이용하여 한국형 MEU모형을 설계한 결과, 기존의 북미-유럽형 모형과 차이가 많이 나는 것으로 나타났으며 성능면에서도 선형로지스틱회귀모형보다 전 산업분류에 걸쳐 한국형 MEU모형이 매우 우수한 것으로 나타났다.

주요용어: 기대효용최대화, 선형로지스틱회귀모형, 신용등급, 파산확률, 커널함수.

1. 서론

기업의 신용 가치는 자본시장에서 많은 영향력을 갖는다. 더욱이 금융기관이 대출을 신청한 기업의 이자율을 산정하는 데는 기업의 신용도가 무엇보다도 중요하다. 일반적으로 이자율 산정 단계를 살펴보면 선형로지스틱모형으로 파산확률 (probability of default)을 추정하고 이를 등급화 (scoring)하여 등급별 이자율을 산정한다. 따라서 신뢰성 있는 파산확률을 추정하는 것이 중요하다고 할 수 있다. Bielecki와 Rutkowski (2002)를 포함하여 파산확률에 대한 논문은 방대하다. 그 중에서 Friedman과 Sandow (2003a)는 확률모형의 표준적인 접근법들 (예를 들어, Vapnik, 1999; Berger, 1985; Hastie 등, 2001)이 종종 위험함수 (risk function)나 손실함수 (loss function)를 구조화하는 방법에 대한 지침이 없음을 지적하고 이를 보완하고자 최대 엔트로피 (maximum entropy)와 효용이론 (utility theory)을 결합한 기대효용최대화 (maximum expected utility) 접근법을 제안하였다.

이항반응변수를 가진 범주형 자료의 판별함수로는 선형로지스틱회귀모형을 사용한다. 그러나 이 모형은 모형의 선형제약으로 인해 반응변수가 강한 단조 함수의 형태를 보이므로

* 본 연구는 고려대학교 특별연구비 (과제번호 KO617001)에 의해 수행되었음.

1) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 정경대학 통계학과, 교수

E-mail: yspark@korea.ac.kr

2) (150-705) 서울시 영등포구 여의도동 27-3, 한국기업데이터, 리스크솔루션 팀

E-mail: jhsong@kedkorea.com

3) (136-701) 서울시 성북구 안암동 5가 1, 고려대학교 통계연구소, 연구조교수

E-mail: cbskust@korea.ac.kr

데이터의 특징을 유연하게 반영하지 못하는 약점을 가지고 있다. Friedman 등 (2003)은 이와같은 선형로지스틱회귀모형의 단점을 보완하고 보다 강력한 통계적 방법론 (statistical learning theory)에 기반한 MEU (Maximum Expected Utility) 모형을 제안하였다. MEU 모형은 선형 형식 (Linear Features), 이차 형식 (Quadratic Features), 커널 형식 (Kernel Features)을 사용하여 데이터의 특징을 보다 유연하게 반영할 수 있는 모형이다 (Friedman과 Sandow, 2003a). 그리고 Cangemi 등 (2003)의 연구에서는 프랑스 도소매업자료를 이용한 선형로지스틱회귀모형과 MEU모형의 성능 평가에서 MEU 모형의 성능이 더 뛰어난 것으로 나타났다. 그러나 이 모형에 포함되어 있는 모수의 일부가 북미와 유럽지역의 자료를 이용하여 경험적 연구에 의해 추정되어져 있는 것이므로 우리나라 기업에 바로 적용하기에는 무리가 있다. 따라서 본 논문에서는 우리나라 중소기업의 자료를 바탕으로 모수를 재추정하여 한국형 MEU모형을 제안하고자 한다.

본 논문에서는 34,057개의 중소기업자료를 이용한 한국형 MEU모형을 설계하여 선형로지스틱회귀모형과 성능을 비교 평가하였으며 전체기업을 제조업, IT기업, 건설업, 도소매업, 그리고 서비스업으로 분류한 후 각각의 모수를 추정하여 산업별 모형을 설계하였고 로지스틱회귀모형과 성능을 비교 평가해 보았다. 현재시점까지의 자료를 이용하여 미래시점에 대한 기업의 평가 (우량 또는 불량)는 파산확률의 실제 응용에 매우 중요하다. 이를 위해 2001년과 2002년 자료를 이용하여 모형을 설계하였고 각 연도별 모형으로 2003년 자료를 판별하여 예측력을 측정하였다.

본 논문의 구성은 다음과 같다. 먼저 2절에서는 선형로지스틱회귀모형과 MEU모형의 방법론에 대해 알아보았고, 3절에서는 우리나라 중소기업 자료를 이용한 한국형 MEU모형을 설계하여 그 성능을 비교평가하였다. 마지막 4절에서는 결론을 정리하였다.

2. 신용평가절차 및 MEU모형

신용평가를 위한 기업의 자료는 파산기업 (default)과 정상기업 (non-default)을 구분하기 위한 이항반응변수와 각종 재무, 거시경제, 시장에 대한 변수를 설명변수로 하는 자료로 구성되어 있는 것이 일반적이다. 본 논문에서는 개별 기업을 관측치라 하고 기업의 파산상태를 나타내는 반응변수를 Y ($Y = 1$: 파산, $Y = 0$: 정상)로 표기하였고 k 개의 설명변수를 $\mathbf{X} = (X_1, \dots, X_k)$ 라 정의하였다. 이 때 설명변수는 자료의 범위가 변수마다 다르고 분포도 일정하지 않기 때문에 동일한 기준으로 모형에 반영할 수가 없다. 따라서 설명변수를 동일한 형태로 만들기 위하여 순위변환 (rank transformation)을 한 후, 0과 1사이의 값으로 변환시켜 모두 같은 분포를 갖도록 하였다 (Skomp 등, 1986; Kane과 Meade, 1998).

2.1. 신용평가절차

MEU모형을 설명하기 전에 일반적으로 사용되는 신용평가절차에 대해 간략하게 살펴보면 다음과 같다. j 번째 개별기업의 파산확률을 p_j (즉, $p_j = \Pr(Y_j = 1)$)라 하고 이 p_j 가 j 번째 관측치에 대한 설명변수의 선형결합 $\mathbf{x}'_j\beta$ 와 다음과 같이 로지스틱 함수에 의해 연계

(link)되어 있다고 가정한다.

$$\log \frac{p_j}{1 - p_j} = \mathbf{x}'_j \boldsymbol{\beta}, \quad j = 1, \dots, n. \quad (2.1)$$

여기서 $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ 는 추정해야 할 모수벡터이다. Y_j 들이 서로 간에 독립인 베르누이 분포를 따른다는 가정 하에 모수 $\boldsymbol{\beta}$ 를 추정하여 j 번째 기업의 파산 확률 \hat{p}_j 을 구하게 된다. 이 때 사용한 모형이 파산확률을 얼마나 정확히 추정했는지 평가하는 도구로서 ROC (Receiver Operator Characteristic) 곡선과 WGRP (Wealth Growth Rate Pickup)가 가장 널리 사용되고 있다 (Friedman과 Sandow, 2003b).

ROC곡선은 '0'인 관측치 중에서 모형에 의해 '0'으로 분류된 관측치의 비율인 특이도와 실제값이 '1'인 관측치 중에서 모형에 의해 '1'로 분류된 관측치의 비율인 민감도에 대한 좌표로 이루어져 있다. 파산확률을 0과 1사이의 특정한 값을 기준으로 나눈 후 분류기준값 이상일 때 '1'로, 그렇지 않으면 '0'의 값을 부여한다. (1-특이도)를 수평축, 민감도를 수직축으로 하여 분류기준값의 변화에 따른 (1-특이도)와 민감도의 값을 그래프에 표시하여 나타낸 곡선이 ROC곡선이다. 일반적으로 곡선아래의 면적값을 구하여 평가에 이용하고 있다. 표 2.1은 Hosmer와 Lemeshow (2000)에서 제시한 ROC곡선의 아래 면적값에 따른 모형평가 기준이다.

표 2.1: ROC곡선의 아래 면적값에 따른 모형평가 기준

ROC곡선의 아래 면적값	평가수준
ROC = 0.5	관별의 의미가 없음 (no discrimination)
0.7 ≤ ROC < 0.8	보통 (acceptable discrimination)
0.8 ≤ ROC < 0.9	좋음 (excellent discrimination)
0.9 ≤ ROC	아주 좋음 (outstanding discrimination)

WGRP는 비교하는 두 모형 각각의 우도함수값의 평균에 대한 차이를 나타낸다.

$$\begin{aligned} \text{WGRP} &= \frac{1}{n} \sum_{j=1}^n [y_j \log(\hat{p}_j) + (1 - y_j) \log(1 - \hat{p}_j)] \\ &\quad - \frac{1}{n} \sum_{j=1}^n [y_j \log(p_0) + (1 - y_j) \log(1 - p_0)], \quad j = 1, \dots, n. \end{aligned}$$

여기서, \hat{p}_j 는 평가하려는 모형에 의해 추정된 파산확률이고 p_0 는 전체 관측치 수 중에서 파산기업 ($Y_j = 1$)의 비율을 나타낸다. 즉, 어떤 정보도 없이 관측된 파산확률을 토대로 판별한 경우와 비교하여 모형을 통한 파산확률의 예측이 우도 함수값 (실제 파산확률과의 유사성)을 얼마나 높일 수 있는가를 알 수 있는 지표이다.

ROC와 WGRP로 모형을 평가하고 그 수준이 타당하다고 판단되는 모형을 선택하여 파산확률을 추정한 후, 이 파산확률에 해당하는 등급을 부여함으로써 기업의 신용등급을 책정한다.

지금까지 신용평가절차를 살펴본 바와 같이 신용평가에서는 얼마나 정확한 파산확률을 측정하는가의 여부가 가장 중요한 사항이다. 그림 2.1은 본 논문에서 사용한 제조업 자료에서 순금융비용부담율을 X_1 으로 하고 유동성현금비율을 X_2 로 하여 각 변수 값에 따른 관측

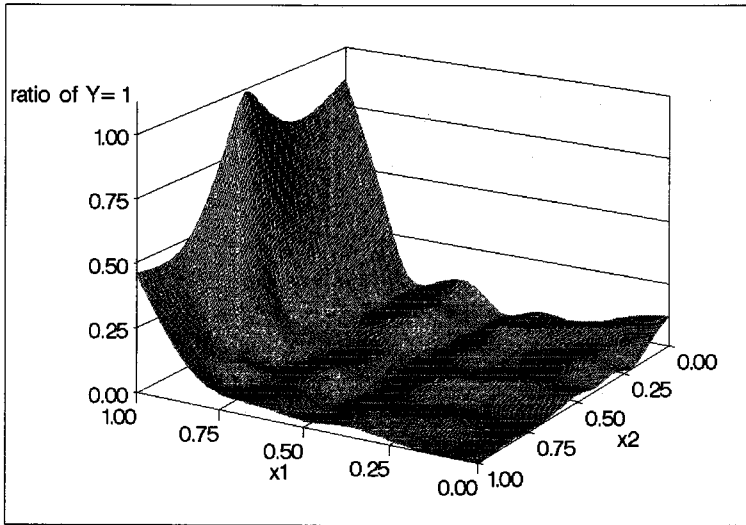


그림 2.1: 실제 자료의 파산확률 대한 평면도

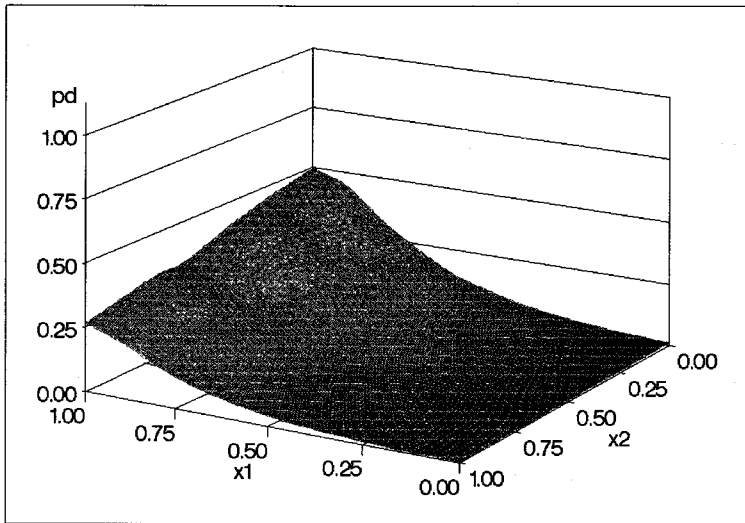


그림 2.2: 로지스틱 모형으로 추정한 파산확률의 평면도

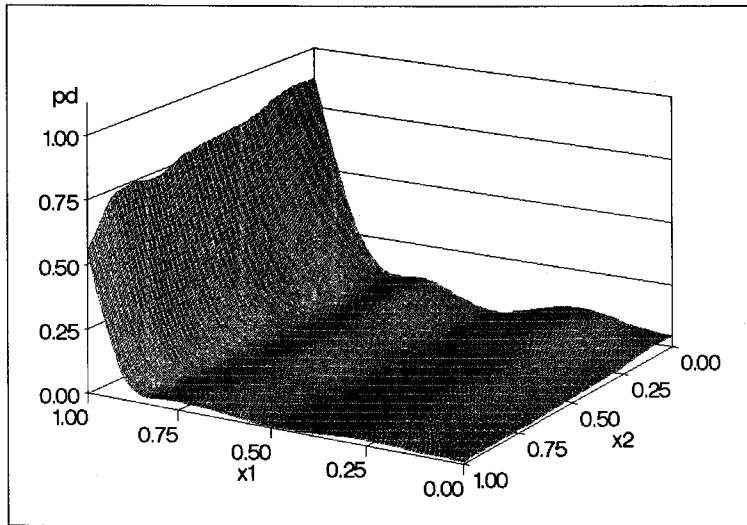


그림 2.3: MEU모형으로 추정한 파산확률의 평면도

된 파산기업의 비율을 평활한 그림이다. 또한 그림 2.2와 그림 2.3은 이 자료를 사용하여 선형로지스틱회귀모형과 2.2절에서 설명될 MEU모형으로 파산확률을 추정하여 평면도를 그린 그림이다. 그림 2.1를 살펴보면 X_1 이 0.75 이상이고 X_2 의 값이 0.25이하인 경우에 파산 비율이 높게 나타나고 있으며, X_1 이 0.25, 0.5, X_2 가 0.0, 0.25 부근에도 파산비율이 높은 돌출된 부분이 있음을 보여주고 있다. 그림 2.2에서와 같이 선형로지스틱모형으로 파산확률을 추정한 경우 그 형태가 평평한 형태가 되어 선형로지스틱회귀모형으로는 그림 2.1의 특성을 반영하는 것이 불가능하게 된다. 그러나, 그림 2.1과 MEU의 모형에 의해 추정된 그림 2.3을 비교해보면 매우 유사한 형태를 보이고 있다는 것을 알 수 있다.

2.2. MEU (Maximum Expected Utility) 모형

Friedman 등 (2003)에서 제안한 MEU방법은 식 (2.1)을 다음과 같이 선형, 이차, 그리고 커널형식으로 파산확률 p_j 를 표현하였다.

$$\log \frac{p_j}{1 - p_j} = \sum_{i=1}^k \beta_{1i} x_{ij} + \sum_{i=1}^k \sum_{m \geq i}^k \beta_{2im} x_{ij} x_{mj} + \sum_{i=1}^k \sum_{l=1}^5 \beta_{3il} \exp \left\{ -\frac{(x_{ij} - a_l)^2}{\sigma^2} \right\}. \quad (2.2)$$

여기에서 $a_l = 0.25l, l = 0, 1, \dots, 4$ 이며 σ 는 복미 및 유럽의 자료를 바탕으로 경험적 연구를 통해 0.35로 고정된 값이다.

식 (2.2)에서의 β 의 추정은 Y_j 가 서로간에 독립인 베루누이 시행이라는 가정에서의 우도 함수로부터 구해진다. 그러나 지나치게 많은 모수의 도입으로 과대적합 문제가 야기될 우

표 2.2: 프랑스 도매업 자료에 대한 모형별 성능 비교

모형	WGRP	ROC
선형로지스틱회귀모형	0.0149	0.838
MEU모형	0.0211	0.843

려가 있어, Friedman 등 (2003)은 손실함수 $F(\beta)$ 를 도입하여 다음과 같은 효용함수 (utility function)를 정의하였다.

$$\text{조정된 } L = \sum_j [y_j \log(p_j) + (1 - y_j) \log(1 - p_j)] - F(\beta). \quad (2.3)$$

여기서 손실함수 $F(\beta)$ 는 과대적합 문제를 완화시켜주는 역할을 하며 $\alpha \sum_{i=1}^k (|\beta_{1i}| + \sum_{m \geq i} |\beta_{2im}| + \sum_{l=1}^5 |\beta_{3il}|)$ 또는 $\alpha \sqrt{\sum_{i=1}^k (\beta_{1i}^2 + \sum_{m \geq i} \beta_{2im}^2 + \sum_{l=1}^5 \beta_{3il}^2)}$ 의 형태를 따른다. α 는 Friedman과 Sandow (2003a)이 제안한 방법을 Park과 Kim (2005)이 재해석하여 도출한 값으로써 자유도가 u 인 카이제곱분포 $\chi_u^2(v)$ 의 cdf 의 역함수 값과 '0' 사이의 값 중에서 효용함수를 최대화 하는 값을 갖는다. 여기서 u 는 β 의 갯수이고 v 는 주어진 신뢰수준이다.

Cangemi 등 (2003)은 프랑스 사기업 중에서 도매업에 대한 1997년부터 2000년까지의 자료를 사용하여 선형로지스틱회귀모형과 MEU모형 (2.2)의 성능을 표 2.2와 같이 비교 평가하였다. 자료는 약 56,000개이고 14개의 재무, 거시경제, 시장에 대한 변수를 사용하였다. WGRP는 선형로지스틱회귀모형보다 0.0062가 증가하였고 ROC는 0.005가 증가하여 선형로지스틱모형보다는 MEU모형의 성능이 더 좋다고 평가할 수 있다. 이처럼 Cangemi 등 (2003)에서는 MEU모형의 우수성을 나타내고 있지만 이는 북미 및 유럽지역의 자료를 바탕으로 σ 와 a_i 를 추정하여 모형 (2.2)와 같이 업종에 관계없이 고정값으로 사용하였다. 따라서 이 고정값을 포함한 모형을 우리나라 기업에 적용할 경우 그 타당성이 의문시 된다. 그리고 경기변동이나 산업구조의 변경 등에 의하여 경제 상황이 변할 수 있다. 그러므로 우리나라 기업의 경제상황에 적합한 σ 를 추정하고 최적의 a_i 값을 선택하여 한국형 MEU모형을 설계해 보고자 한다.

3. 우리나라 중소기업 자료를 이용한 사례분석

우리나라에서는 자산규모 70억 이상인 기업을 외부감사 대상으로 선정한다. 본 논문에서는 2001년부터 2003년까지 외부감사 대상에 속하지 않은 기업 중 총 자산규모 20억 이상인 34,057개의 중소기업을 대상으로 모형을 수립하고, 유럽형 MEU모형의 모수와 한국형 MEU모형의 모수에 대한 차이 여부를 알아보려고 한다. 또한, 각 업종별로 모형을 세분화하여 업종에 따라 모형의 차이가 있는지 살펴보고자 한다. 기업의 신용등급은 현재의 자료를 바탕으로 미래에 파산할 확률에 의해 결정된다. 그러므로 2001년과 2002년 자료를 가지고 모형을 설계하였고 각 연도별 모형으로 2003년 자료를 판별하여 예측력을 측정하였다.

업종은 제조업의 16,941개의 중소기업, 3,612개의 IT산업, 5,900개의 건설업, 5,137개의 도소매업, 2,467개의 서비스업으로 분류하였고, 연체일수가 90일 이상인 기업을 파산기업으로 분류하였다. 분석을 위하여 SAS의 행렬연산 도구인 SAS/IML 모듈을 이용하여 프로그래밍 하였다.

3.1. 변수선택 및 모형설계

사례분석을 위하여 우리나라 중소기업의 81개 재무비율을 수집하고 성장성, 수익성, 현금상환능력, 안정성, 활동성, 추세-변동성을 나타내는 6개의 특성집단로 각각 분류한 후 (김민환, 2000), 자료를 파산기업인 불량집단 과 파산하지 않은 정상집단으로 분리하여 두 집단의 차이를 잘 나타내 주는 변수를 t -검정과 윌콕슨 검정 (Wilcoxon-test)을 이용하여 선택하였다. 업종별로 선택된 변수들은 표 3.1에 정리 되어있다.

식 (2.3)을 최대화시키는 모수를 찾는 방법으로는 뉴턴-랩슨 (Newton-Raphson) 방법을 사용하였다. 유럽형 MEU모형에서는 $a_1 = 0, a_2 = 0.25, a_3 = 0.5, a_4 = 0.75, a_5 = 1$ 로 지정하였으나 반드시 5가지가 모두 유의하다고는 볼 수 없으므로 최적의 a_i 을 찾기위해 다음과 같이 a_i 과 σ 를 구하였다.

- 1단계: $a = \{0, 0.25, 0.5, 0.75, 1\}$ 를 갖는 모형에서 초기 σ 를 추정한다.

표 3.1: 최종변수 선택

업종	선택된 변수
전체	총자본경상이익율, 매출액영업이익율, 순금융비용부담율, 차입금의존도, 자기자본비율, 평균차입금대매출액, 유동성현금비율, cash/current assets, 총자산회전율, 매출액변동계수, 매입채무회전기간
제조업	총자본경상이익율, 매출액영업이익율, 금융비용부담율, 순금융비용부담율, 차입금의존도, 자기자본비율, 평균차입금대매출액, 유동성현금비율, EBITDA대금융비용 (영업이익)
IT기업	매출액영업이익율, 차입금의존도, 평균차입금대매출액, 유동성현금비율, 총자산회전율, 고정자산회전율, 영업이익추세, 현금흐름추세
건설업	차입금의존도, 평균차입금대매출액, 유동성현금비율, 총자산회전율, 영업자산회전율, 영업이익추세, 자산가치추세, 금융비용 대 부채비율, 영업이익대비 총자산비율
도소매업	순금융비용부담율, 당좌비율, 자기자본비율, 유동성현금비율, Interest Cover, 매출액변동계수, 단기부채상환능력
서비스업	총자본경상이익율, 순금융비용부담율, 당좌비율, 유동성현금비율, 자산가치추세

- 2단계: 추정된 σ 를 모형에 포함시키고 $a = \{0, 0.25, 0.5, 0.75, 1\}$ 에서 각각의 값을 하나씩 제거하여 가장 큰 목적함수 식 (2.3)의 값을 갖는 모형을 선택한다.
- 3단계: 선택된 a 를 포함한 모형에서 다시 σ 를 추정한다.
- 4단계: 식 (2.3)이 최대가 될 때까지 2단계와 3단계를 반복한다.

이와같이 네단계를 통해 추정된 모수는 표 3.2에 업종별로 정리되어 있다.

표 3.2: 최종 모형에 사용되는 모수

업종	σ	a
전체 산업	0.019	{0, 0.25, 0.5, 1}
제조업	0.241	{0, 0.5, 0.75, 1}
IT산업	0.098	{0, 0.25, 0.5, 0.75, 1}
건설업	0.012	{0, 0.25, 0.5, 1}
도소매업	0.249	{0, 0.25, 0.5, 1}
서비스업	0.024	{0, 0.25, 0.5, 0.75, 1}

유럽형 MEU모형 식 (2.2)의 $\sigma = 0.35$, $a_1 = 0$, $a_2 = 0.25$, $a_3 = 0.5$, $a_4 = 0.75$, $a_5 = 1$ 와 비교할 때, 표 3.2의 모수는 분명한 차이를 보이고 있다. 특히, 모든 산업에 걸쳐, σ 가 북미·유럽형의 $\sigma = 0.35$ 보다 작게 나온 것은, 한국형이 북미·유럽형보다 돌출부분이 더 뾰족하다는 (즉, 파산확률이 몰려있는) 것을 의미한다. 건설업의 경우, 순위변환을 통한 재무변수의 값이 0, 0.25, 0.5, 1를 중심으로 가장 뾰족한 돌출형패를 보여주고 있다. 전체산업을 기준으로 볼 때 재무변수의 값 0.75에는 파산확률의 돌출부분이 없는 것으로 나타났으나, 제조업, IT산업, 그리고 서비스업의 경우, $a = 0.75$ 에 파산확률이 높게 나타나고 있다.

3.2. 모형평가

본 절에서는 표 3.2의 선택된 모형이 선형로지스틱회귀모형과 비교하여 그 성능이 얼마나 향상되었는지를 알아보려고 한다. 성능평가도구로는 ROC와 WGRP를 사용하였고 전체자료와 각각의 업종별 자료에 대해 모형설계 집단 (training set)과 평가집단 (test set)을 임의의 표본에 의해 8:2로 나누어 30회 반복하여 측정하였다.

표 3.3의 모형평가결과에서는 전반적으로 MEU모형이 선형로지스틱회귀모형보다 그 성능 뛰어남을 보여주고 있다. 특히, 제조업의 경우에는 ROC가 0.08이 증가하여 가장 큰 폭의 증가를 보여주고 있다.

기업의 신용등급의 정확도는 현재시점까지의 해당 재무자료를 이용하여 미래시점의 파산확률을 얼마나 정확하게 예측하느냐에 달려있다. 이를 위해 자료를 연도별로 나누어 2001년과 2002년 자료로 모형을 설계하고 (즉, in-sample data) 이 모형을 2003년 자료에 적용시켜 (out-of-sample data) 예측력을 측정하였다.

표 3.3: 업종별 자료의 모형평가

업종	로지스틱모형		MEU모형	
	ROC	WGRP	ROC	WGRP
전체 산업	0.805 (0.009)	0.050 (0.004)	0.874 (0.008)	0.097 (0.005)
제조업	0.800 (0.014)	0.037 (0.005)	0.883 (0.015)	0.094 (0.010)
IT산업	0.864 (0.015)	0.138 (0.015)	0.873 (0.017)	0.157 (0.022)
건설업	0.771 (0.025)	0.036 (0.009)	0.789 (0.020)	0.043 (0.012)
도소매업	0.792 (0.025)	0.028 (0.009)	0.813 (0.027)	0.038 (0.012)
서비스업	0.840 (0.043)	0.041 (0.016)	0.875 (0.042)	0.051 (0.036)

*괄호 안의 값은 표준편차

표 3.4: 연도별 자료의 모형평가

in-sample data	로지스틱모형		MEU모형	
	ROC	WGRP	ROC	WGRP
2002년	0.807	0.007	0.855	0.018
2001년	0.806	0.013	0.850	0.025
2001년과 2002년	0.808	0.010	0.855	0.026

표 3.4는 2003년도 파산확률을 예측하는데 있어서 전년도 자료(2002년 자료)나 2년전의 자료(2001년 자료) 모두 비슷한 예측력을 보여주고 있으며 2개년도(2001년과 2002년) 자료 모두를 이용하는 것이 모형에 관계없이 가장 우수한 것으로 나타났다. 또한, MEU모형이 모형설계자료에 관계없이 선형로지스틱모형보다 우수한 것으로 나타나고 있다.

표 3.5는 2.1절에서 언급한 ROC곡선의 민감도와 특이도, 정분류된 자료의 비율을 나타내는 정분류율이 동일한 값을 갖는 지점을 분류기준값으로 선정한 결과이다. 각기업의 파산확률이 이 값보다 크면 불량기업으로 분류하고 그렇지 않으면 우량기업으로 분류한다. 이 값을 기준으로 산업별로 선형로지스틱모형에서 우량기업을 불량기업으로, 불량기업을 우량기업으로 오분류한 관측기업 가운데, MEU모형에 의해 정분류한 관측기업체의 수를 표 3.6에 정리하였다. 표 3.6에 따르면 선형로지스틱모형은 총 34,057개의 중소기업 가운데 우량인 9,056개의 기업을 불량으로 오분류하였고 불량인 644개의 기업을 우량으로 잘못 분류하였다. 그러나 MEU모형을 적용하면 선형로지스틱모형에 의해 우량→불량으로 오분류되었던 9,056개의 중소기업중 2,132개의 기업이 정분류되었고 불량→우량으로 오분류된 644개의 중소기업중 86개의 중소기업이 정분류되었다. 은행의 여신업무 입장에서 우량→불량으로의 오분류보다는 불량→우량의 오분류가 매우 큰 손실을 초래한다. 이러한 관점에서 볼 때, MEU 모형을 전체산업 보다 업종별로 적용하는 것이 로지스틱선형모형에 비해 훨씬 효율적임을 알 수 있다.

표 3.5: 업종별 자료의 분류기준값

업종	로지스틱모형	MEU모형
전체 산업	0.073	0.062
제조업	0.069	0.050
IT산업	0.147	0.128
건설업	0.083	0.086
도소매업	0.060	0.052
서비스업	0.051	0.040

표 3.6: 로지스틱모형 대비 MEU모형의 업종별 정분류 증가량

업종	로지스틱모형의 오분류		MEU모형의 정분류 증가량	
	우량→불량	불량→우량	우량기업	불량기업
전체 산업 9056	644	2132	86	
제조업	3390	375	153	153
IT산업	565	140	34	39
건설업	1673	124	134	16
도소매업	1623	53	206	0
서비스업	474	33	14	14

4. 결론

본 논문에서는 우리나라 중소기업자료를 이용하여 업종별 한국형 MEU모형을 제안하였다. 기존에 제안된 MEU모형은 북미 및 유럽지역의 자료를 바탕으로 커널형식에 포함되어 있는 모수를 추정하여 사용하였고 그 효율성을 보여 주었다. 그러나 이 추정된 모수는 우리나라 기업의 특징을 반영한 모수라고 볼 수 없으므로 우리나라 중소기업자료를 이용하여 모수를 재추정하여 한국형 MEU모형을 제안하였다.

이렇게 설계된 한국형 MEU모형과 기존의 선형로지스틱회귀모형의 성능을 비교평가한 결과, 먼저 한국형 MEU모형은 북미 및 유럽형 MEU모형과 많은 차이를 보여주었고 한국형 MEU모형이 선형로지스틱모형보다 ROC와 WGRP에서 더 높은 값을 나타냈다. 특히, MEU모형은 선형로지스틱모형으로부터 관찰된 불량→우량의 오분류율을 현저히 감소시키고 있음을 알 수 있었다. 연도별로 자료를 나누어 2001년과 2002년 자료는 모형을 설계하고 각 연도별 모형으로 2003년 자료를 판별하여 예측력을 비교한 결과, 제시된 한국형 MEU모형이 기존의 선형로지스틱모형의 예측력보다 모형에 사용한 자료의 연도수에 관계 없이 우수하다는 것을 보여 주었다.

참고문헌

- 김민환 (2000). <재무관리>, 도서출판 흥, 서울.
- Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer-Verlag, New York.
- Bielecki, T. R. and Rutkowski. M. (2002). *Credit Risk: Modeling, Valuation and Hedging*, Springer-Verlag, Berlin.
- Cangemi, B., Servigny, A. D. and Fiedman, C. (2003). Standard & poor's credit risk tracker for private firms, *Standard & Poor's Working Paper*.
- Friedman, C. and Sandow, S. (2003a). Learning probabilistic models: an expected utility maximization approach, *The Journal of Machine Learning Research*, 4, 257-291.
- Friedman, C. and Sandow, S. (2003b). Model performance measures for expected utility maximizing investors, *International Journal of Theoretical and Applied Finance*, 6, 355-401.
- Friedman, C., Huang, J., Servigny, A. D. and Salinas, E. (2003). A utility-based private firm default probability model, *Standard & Poor's Working paper*.
- Hastie, T., Tibshirani, R. and Friedman, J. H. (2001). *The Elements of Statistical Learning*, Springer-Verlag, New York.
- Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression*, 2nd ed., John Wiley & Sons, New York.
- Kane, G. and Meade, N. (1998). Ratio analysis using rank transformation, *Review of Quantitative Finance and Accounting*, 10, 59-74.
- Park, Y. S. and Kim, K. W. (2005). Report on credit risk tracker, *Working paper*.
- Skomp, S. E., Cronan, T. P. and Seaver, W. L. (1986). On application of the rank transformation discrimination method to financial problems, *The Financial Review*, 21, 473-483.
- Vapnik, V. N. (1999). *The Nature of Statistical Learning Theory*, 2nd ed., Springer-Verlag, New York.

[2007년 6월 접수, 2007년 8월 채택]

Korea-specified Maximum Expected Utility Model for the Probability of Default*

You Sung Park¹⁾ Ji Hyun Song²⁾ Bo-Seung Choi³⁾

ABSTRACT

A well estimated probability of default is most important for constructing a good credit scoring process. The maximum expected utility (MEU) model has been suggested as an alternative of the traditional logistic regression model. Because the MEU model has been constructed using financial data arising from North America and European countries, the MEU model may not be suitable to Korean private firms. Thus, we propose a Korea-specific MEU model by estimating the parameters involved in kernel functions. This Korea-specific MEU model is illustrated using 34,057 private firms to show the performance of the MEU model relative to the usual logistic regression model.

Keywords: Maximum expected utility, linear logistic regression, credit scoring process, probability of default, kernel function.

* This research was supported by the research fund (KO617001) of Korea University.

1) Professor, Department of Statistics, Korea University, 1 Anam-Dong, Sungbuk-Gu, Seoul 136-701, Korea

E-mail: yspark@korea.ac.kr

2) Risk solutions Team, Korea Enterprise Data Co., Ltd., 27-3 Yeouido-Dong, Yeongdeungpo-Gu, Seoul 150-705, Korea

E-mail: jhsong@kedkorea.com

3) Research Associate professor, Institute of Statistics, Korea University, 1 Anam-Dong, Sungbuk-Gu, Seoul 136-701, Korea

E-mail: cbskust@korea.ac.kr