

## 범주형 재무자료에 대한 신용평가모형 검증 비교

홍종선<sup>1)</sup>, 이창혁<sup>2)</sup>, 김지훈<sup>3)</sup>

### 요약

재무자료에 대한 신용평가모형은 각각의 재무변수를 평활한 예측부도율로 변환하여 사용한다. 본 연구에서는 연속형 재무자료를 변환하여 설정된 신용평가모형의 문제점을 살펴보고, 연속형 재무변수를 다양한 형태로 범주화한 신용평가모형들을 제안한다. 범주형 재무자료를 사용해서 개발한 여러 종류의 신용평가모형들의 성과를 다양한 적합성 검증 방법으로 비교하고, 범주형 재무자료를 이용한 신용평가모형의 유용성을 토론한다.

주요용어: 단계별 변수선택; 신용평가모형; 재무자료; 범주형 자료; 성과; 적합성 검증.

### 1. 서론

금융회사의 수익 창출을 위한 모든 영업활동에는 필연적으로 리스크를 포함하고 있다. 리스크를 적절히 수용하고 관리하면서 적정한 수익을 확보해 나가는 과정이야말로 금융회사 업무의 본질인 동시에 핵심이라고 할 수 있다. 이러한 관점에서 신용리스크는 금융시장에서 가장 오래된 형태의 리스크이며, 은행을 비롯한 금융회사 입장에서 가장 중요한 리스크로 간주되어 왔다. 신용리스크는 채무불이행 위험뿐만 아니라 채무자(또는 보유자산)의 신용도(credit quality)가 하락할 때, 예를 들어 신용등급(credit rating)이 하향 조정될 때 자산이나 계약의 시장가치가 하락하여 발생할 수 있는 손실위험까지 포함하는 개념으로 정의된다 (검사지원국, 2004).

금융기관은 신용리스크를 관리함에 있어 엄격한 신용평가를 통해 거래상대방을 신중히 선별하는 한편, 신용리스크가 특정 부분에 집중되는 것을 억제하고, 대출매각, 증권화 등을 통해 이미 부담하고 있는 신용리스크를 해소하는 등의 방법을 통해 신용리스크를 관리하고 있다. 신용리스크 관리기법들 중에서 기업 등 개별차주의 재무 및 비재무 상황을 종합적으로 평가하여 그 결과에 따라 일정 수준의 신용등급을 부여하는 모형을 신용평가모형이라 한다. 일반적으로 금융기관에서 기업체를 대상으로 사용하는 신용평가모형은 전문가 판단모형과 재무정보에 기반을 둔 신용평가모형이 결합한 형태이다. 전통적인 금융기관 신용평가방식은 고도의 훈련된 대출담당자의 직관에 의존하는 전문가시스템(expert system)을 중심으로 발전되었으며, 현재도 신용의사결정과정에서 중요하게 활용되고 있

1) (110-745) 서울, 종로구 명륜동 3-53, 성균관대학교 경제학부 통계학전공, 교수.

교신저자: cshong@skku.ac.kr

2) (110-745) 서울, 종로구 명륜동 3-53, 성균관대학교 응용통계연구소, 연구원.

3) (110-745) 서울, 종로구 명륜동 3-53, 성균관대학교 응용통계연구소, 연구원.

다. 그러나 전문가의 경험과 주관에 따라 상이한 평가가 이루어짐에 따라 객관성 및 일관성 유지의 문제가 발생함에 따라 이를 통제하기 위해 재무정보를 이용한 방법과 결합하여 활용된다.

재무정보를 이용한 신용평가모형은 거래상대방이 일정기간 내에 부도(default) 또는 부실화될 가능성을 수치화된 자료를 이용하여 점수화함으로써 평가자의 주관적인 판단을 배제한 객관적 결론을 유도하여 산출하는 것이 목표이다. 따라서 과거 일정기간 동안 실제 부도난 기업들의 재무자료를 토대로 통계적인 분석을 수행하여 유의한 것으로 밝혀진 설명 변수들을 추출하고 이 변수들을 이용한 부도확률함수를 제시함으로써 기업의 잠재적 부도 가능성을 계량화하는 방법이 재무정보에 근거한 신용평가모형이다.

신용등급(credit rating)을 산출하기 위한 통계적 방법 그리고 신용평가(credit assessment)에 관하여는 Hand와 Henley (1997)를 비롯한 많은 연구가 진행되었는데 본 논문에서는 비재무자료를 제외한 재무자료만을 사용한 신용평가모형(credit rating model, credit evaluation model)에 대해 연구한다. 신용평가모형은 대부분 연속형 재무자료(financial data)를 수집하여 재무비율들을 생성한 후, 각각의 변수를 연도별 변동을 줄이면서 평활하게 변수변환한다. 그리고 변수를 선택하여 최적의 모형을 개발하는데, 특정 분야의 전문가들만이 다룰 수 있는 전문적 지식습득을 기초로 하는 분야이다. 본 연구에서는 기존 신용평가모형의 문제점을 토론하고, 연속형 재무자료를 이산화해서 범주형 재무자료를 이용한 신용평가모형을 제안한다. 그리고 연속형 재무자료의 신용평가모형과 분류성과(performance) ‘평균차이’(D), ‘불량 누적비율 50%에 대한 우량의 누적비율’(1-PH), ‘일치성’(Concordant), ‘콜모고로브-스미노브’(K-S), ‘accuracy ratio’(AR), ‘area under receiver operating characteristic’(AUROC) 등의 통계량들 (박병수, 2004; 임종건, 2005; Engelmann 등, 2003; Joseph, 2005; Stein, 2005; Thomas 등, 2004, 등 참조)을 사용하면서 비교 평가하고 그 결과를 살펴보면서 범주형 재무자료의 유용성을 토론 한다.

본 논문의 구성은 다음과 같다. 2절에서는 국내의 특정한 금융기관에서 개발한 신용평가모형 설정과정을 설명한다. 3절에서는 현재 사용하고 있는 신용평가모형의 문제점을 지적한 후에, 각각의 재무 변수를 범주형 변수로 변환하는 방법을 대안적으로 제시한다. 그리고 변환된 범주형 자료를 신용평가모형에 적용해 연속형 재무자료를 이용한 기존의 모형과 비교한다. 일반적인 신용평가모형 개발과정에서 일변량 검정과 상관분석 등을 이용한 여러 단계의 변수선택을 하면서도 모형의 설명력, 직관성, 안정성 등을 고려하여 추가하거나 대체된 변수들을 사용한다. 4절에서는 통계적인 방법인 단계별 변수선택법을 이용해 새로운 변수들로 구성된 신용평가모형을 제안하고, 새로 선정된 재무 변수들을 범주화하여 단계별 선택법을 이용한 신용평가모형 분류성과 결과와 비교한다. 마지막으로 5절에서는 본 연구에서 제안한 신용평가모형에 대하여 결론을 내린다.

## 2. 신용평가모형 개발과정

본 논문이 참고하고 있는 신용평가모형을 개발하는 과정은 방대한 재무자료를 수집한 다음에 변수생성, 변수변환, 1·2차 변수선택 그리고 최종 모형 선택과정을 통해 개발이 이

루어진다.

첫번째 자료수집 및 변수생성 단계에서는 재무모형 개발에 적용 가능한 자료를 이용하여 각 재무비율을 변수로 생성하고, 두번째 변수변환 단계에서는 각 재무변수의 연도별 자료를 Loess 방법을 이용하여 평활(smoothing)한 예측부도율로 변환한다. 그리고 세번째 단계에서는 예측부도율로 변환한 자료에 대한 신용평가모형의 Concordant, AR 통계량 등을 통해 변수의 유의성을 살펴보면서 1차 변수선택을 하고, 변수들의 상관관계를 통해 2차 변수선택을 실시한다. 그리고 마지막으로 네번째 최종모형 선택 단계에서는 로지스틱 회귀분석을 통한 적합한 모형성과 비교와 현업 및 추가적인 모형조합의 성과를 비교하기 위하여 여러 종류의 적합성 검증 방법(D, 1-PH, K-S, AR, AUROC, Concordant 통계량)을 통해 비교함으로써 최종적인 신용평가모형을 결정한다.

## 2.1. 자료수집 및 변수생성

자료수집은 1994년부터 2005년까지 외감기업 중 매출액 1000억 이상의 대기업들에 관한 연도에 따른 4268건(정상: 4101건, 부도: 167건)의 재무, 불량, 감사의견, 법정관리 및 회회정보 자료를 사용하였으며, 업종구분은 가변수(dummy variable)로 제조업, 건설업, 서비스업, 도소매업, IT 등 5개 업종으로 구분하였다. 또한 은행연합회 불량 자료와 워크아웃 자료로 부도를 정의하고 재무제표별 관측기간 내 부도가 발생할 경우는 불량(부도기업)으로, 그 외 재무제표별 관측기간 내 부도가 발생하지 않은 경우에는 우량(정상기업)으로 정의하였다.

자료의 변수들은 재무모형 개발을 위해 모 기업평가회사의 기업 데이터베이스에서 관리되는 재무분석에 영향력 있는 변수들과 이미 다른 모형개발에서 검증되고 있는 변수들을 추가하여 182개 변수의 재무평가항목을 선정하였다.

## 2.2. 변수변환

변수변환 과정의 첫번째는 'By year method' 단계로 우선 182개의 재무변수 각각을 연도별로 재무비율의 구성이 2%씩 되도록 50개의 구간(bucket)으로 나눈다. 그리고 각 연도에 해당하는 2%, 4%, ..., 100%의 구간별 부도 빈도수를 계산한 뒤 50개 동일 구간 내 연도별 자료를 묶어 각 구간을 하나의 부도율을 계산한다. 두번째는 비모수적 평활(non-parametric smoothing) 단계로 재무변수의 각 백분위수(percentile)에 해당하는 구간별 부도율은 변동이 심하므로 평활한(smoothing) 추세를 나타내는 Loess 방법(해당위치의 값을 주위 값으로 추정된 선형식 값으로 대체하여 주변의 값은 비슷한 값을 가지도록 하여 위치에 따른 변동을 줄이는 방법으로 자세한 설명은 Cleveland (1979)와 Cohen (1999)의 논문을 참조)을 이용하여 구간별 예측부도율로 변환한다. 그림 2.1은 Loess 방법을 이용한 평활곡선(smoothing curve)을 예로 든 것이다. 그림 2.1의 곡선그림은 각 재무비율 값(X축)을 크기순으로 정렬하여 50개의 구간으로 나누고, 각 구간별 실제 부도율(Y축)을 계산하여, 구해진 비율의 부도율 값인 50개 구간에 해당하는 위치를 Loess 방법으로 loess 모수(=0.4, 0.5, 0.6)에 따라 평활한 것이다. 그림 2.1에서 곡선그림은 예측부도율의 역전이 일어나지 않도록 평활 정도를 크게 하기 위하여 loess 모수를 0.6으로 결정하였다.

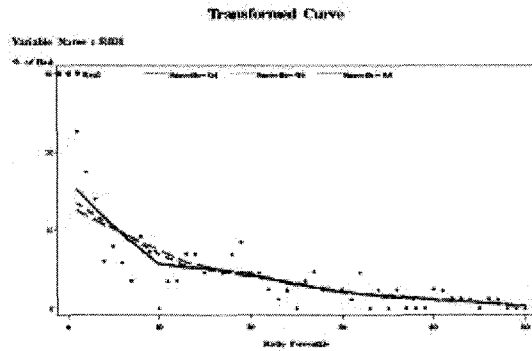


그림 2.1: Loess방법을 이용한 평활곡선

표 2.1: 선택된 변수

범주항목	1차 변수선택	2차 변수선택	
		방법1 결과	방법2 결과
규모(size)	4	1	
안정성	Leverage	10	3
	Retained Earning	5	1
부채상환능력 (Debt Coverage)	10	2	5
수익성(Profitability)	6	2	
유동성(Liquidity)	4	2	
성장성(Growth)	4	1	
활동성(Activity)	7	1	1
전체	50	12	10

### 2.3. 변수선택

변수선택은 두 단계로 나누어지는데, 우선 각 모형별 182개 각각의 재무변수를 변환한 자료에 대한 신용평가모형의 검증 통계량( $D$ ,  $1-PH$ , Concordant,  $K-S$ , AR, AUROC)을 살펴보면서 유의한 변수를 선택한다(1차 변수선택). 일반적으로 변수선택 기준은 각 범주항목 내에서 AR, AUROC, Concordant 값이 높고,  $K-S$ 통계량 값이 20% 이상이거나 Loess 그래프에서 비율의 부호(방향)가 적절한 비율을 선택하는데, 만약 유사한 의미의 비율인 경우는 AR이 높거나 계산식이 간명하고, 업계 경험상 선호되는 비율을 선택한다. 표 2.1의 1차 변수선택 칸은 변수선택 기준에 의해 1차로 선택된 변수들이다.

2차 변수선택은 1차 변수선택에 의해 선정된 변수들을 또 다른 방법을 적용시켜 선택하는 과정으로, 각 모형별로 상관분석을 통해 다음 두 가지 방법을 이용하여 2차 변수선택한다. 첫번째 방법은 최다의 범주항목을 살리면서 각 범주항목 내 재무변수의 상관분석을 통해 선택하거나 같은 범주항목 내에서 상관관계가 높은 변수들 중 AR이 낮은 것을 제외

표 2.2: 최종 신용평가모형 결과와 분류성과

범주항목	변수	Estimate	Wald Chi-Square	p-값		
	상수항	-7.2094	420.5131	<.0001	검증 통계량	분류성과
	건설업	-0.7051	6.1315	0.0133		
	서비스업	-1.1910	5.7442	0.0165		
	IT	0.5629	4.1977	0.0405		
안정성	(부채총계-이익 잉여금)/(유형자산)	0.1967	22.5573	<.0001	<i>D</i>	1.4035
	차입금/자기자본	0.1155	20.1821	<.0001	1-PH	0.9722
부채상환능력	총C/F대차입금 비율	0.0768	8.5625	0.0034	<i>K-S</i>	0.6676
	경상이익이자 보상비율	0.0906	6.4049	0.0114	AR	0.7819
유동성	단기유동성비율	0.0581	4.3810	0.0363	AUROC	0.8909
합동성	유동부채/매출액	0.0960	13.5367	0.0002	Concordant	0.8830

하고 2~3개의 변수를 선택한다. 만약 유사한 의미의 변수인 경우에는 AR이 높거나 계산식이 간명하고 업계경험상 선호되는 변수를 선택한다. 두번째 방법은 범주항목에 상관없이 AR이 높은 순으로 재무변수의 상관분석을 통해 선택하거나 상관관계가 높은 변수들 중 AR이 높은 변수를 선택한다. 결과적으로 표 2.1의 2차 변수선택 결과를 보면 방법1에서는 12개의 변수가 선택되었고, 방법2에서는 10개의 변수들로 범주항목을 고려하지 않았음에도 어느 정도 범주항목별로 고르게 선택되었음을 알 수 있다.

### 2.4. 최종모형

최종 재무모형 선정은 방법1과 방법2에서 선택된 재무변수들을 이용해 각 조합을 생성하고 모형들의 분류성과를 검증(AR값 중심으로)하고 비교해 분류성과가 높은 결과를 선택하는데 대체로 변수가 6~7개 사이일 때의 조합이 분류성과가 높고 안정적으로 나타났다.

최종적으로 변수가 6~7개 사이일 때의 조합을 대상으로 각 조합별 모형들의 구성 위험요인(risk factor)들이 모형 안에서 통계적으로 유의적인지 확인한 후, 금융전문가들이 모형의 설명력, 선택된 변수의 이해가능성, 직관성, 모형의 안정성 등을 포괄적으로 고려하여, 일부 새로운 변수들을 추가하거나 대체한 후 6개의 재무변수와 3개의 가변수를 이용하여 로지스틱 회귀분석을 실시하고, 모형의 분류성과를 비교하기 위하여 6:4의 비율로 훈련자료(train data)와 타당성 검사자료(validation data)로 분리하여 여러 적합도 검증통계량(*D*, 1-PH, *K-S*, AR, AUROC, Concordant)을 통해 비교함으로써 최종 모형을 결정한다. 선정된 최종 신용평가모형의 회귀계수를 산출해 표 2.2에 나타내었다(일반적으로 최종등급산출은 재무모형(통계모형)과 전문가판단모형인 비재무모형의 결합으로 모형의 성과 및 예측력을 향상시키는데 목적이 있으며, 모형결합을 시행하기 위해 스코어 변환작업을 시행하고 그에 맞는 등급을 산출하는 것이 보편적이다. 그러나 본 연구는 통계적 신뢰

성 확보를 위해 전문가판단모형인 비재무모형을 제외하고 재무모형만을 가지고 모형을 개발함).

표 2.2의 최종재무모형 분류성과의 결과를 살펴보면 재무모형의 신용평가모형에서 4,268건수 중 불량률은 3.91%가 나왔으며, 분류성과를 보았을 때 1-PH, AR, AUROC 그리고 Concordant 통계량들은 매우 큰 값을 갖고 있고 자유도에 의존하는 *D*와 *K-S* 통계량들의 값도 유의한 큰 값을 나타내고 있으므로 모형의 예측률이 높다고 판단할 수 있다.

### 3. 대안적인 신용평가모형

#### 3.1. 기존 신용평가모형의 문제점과 개선 방법

기존의 신용평가모형은 재무평가항목들의 재무자료를 수집해 각 변수마다 Loess 방법을 이용하여 재무변수들을 각 변수에 대응하는 예측부도율로 변환하여 모형을 개발하였다. Loess 방법을 사용하기 위해 'By year method' 단계에서 각 연도별로 구성 비율이 2%가 되게 50개의 구간으로 자료를 나눈 후, 동일 구간 내 연도별 자료를 하나의 부도율로 계산하였다. 'By year method'에 의한 Loess 방법을 사용함으로써 연도별 재무변수의 흐름이 변동함에도 불구하고 그러한 특수상황이 무시됨에 따라 정보의 손실이 예상된다. 그리고 설명변수들인 재무변수 대신에 각 설명변수의 예측부도율로 사용하였기 때문에 다양한 재무변수들의 고유의 정보를 잃어버린다는 문제점을 지적한다.

연속형 변수인 재무자료를 변수변환한 부도율을 사용하고, 분류성과 등 여러 단계를 거쳐 최종모형을 선정하였다. 그러나 기업이 정보유출을 꺼리거나 허위 자료를 배포할 가능성이 클 경우, 정확한 자료를 수집하는데 많은 어려움이 생길 수 있다. 만약 설명변수인 연속형 변수를 소수의 수준을 갖는 이산형 변수로 변환하여 범주형 자료를 이용한 신용평가모형의 결과가 연속형 재무자료를 이용한 신용평가모형과 비슷한 결과를 나타낸다면, 범주형 자료는 모형설계에 있어서 의미있는 자료가 될 것이다. 이것은 비록 정확한 재무자료를 얻지 못할 경우에 개략적인 순위 등과 같은 정보(예를 들어 전체를 네 등분하여 상, 중상, 중하, 하 정도로 구분이 가능한 경우)를 알고 있다면, 수집하기 어려운 여러 재무변수들을 단순하게 범주형으로 변환하여 모형을 개발할 수 있겠다. 그리고 기존의 모형에서는 각 설명변수 대신에 대응하는 예측 반응변수인 예측부도율로 사용하였으나, 범주형으로 변환한 자료를 이용한 신용평가모형에서는 각 재무변수들의 고유의 정보를 유지할 수 있다. 또한 Loess 방법을 이용하지 않기 때문에 연도별 재무자료의 특수상황의 정보를 손실하지 않는다는 장점이 있다.

기존모형에서 사용되고 있는 자료는 대부분이 한쪽으로 치우쳐 있으며, 이러한 자료를 직접 적용할 경우 왜도의 발생으로 평가방법에 영향을 미칠 수 있다. 왜도의 발생을 줄이기 위해 연속형을 범주형으로 변환한다면 왜도로 인해 발생할 수 있는 영향을 좀 더 줄일 수 있다. 그림 3.1은 최종 선택된 6개의 변수에 대한 그래프를 그려본 것이다.

신용평가모형을 개발하는 과정에서, 각 재무자료를 예측부도율로 변환하여 설명변수로 사용하는데, 본 연구에서는 변환 전의 원자료(raw data)의 재무변수를 범주형으로 변환하는 직접적인 변환방법을 사용하여 신용평가모형을 개발하는 방법을 제안한다. 이 방법은

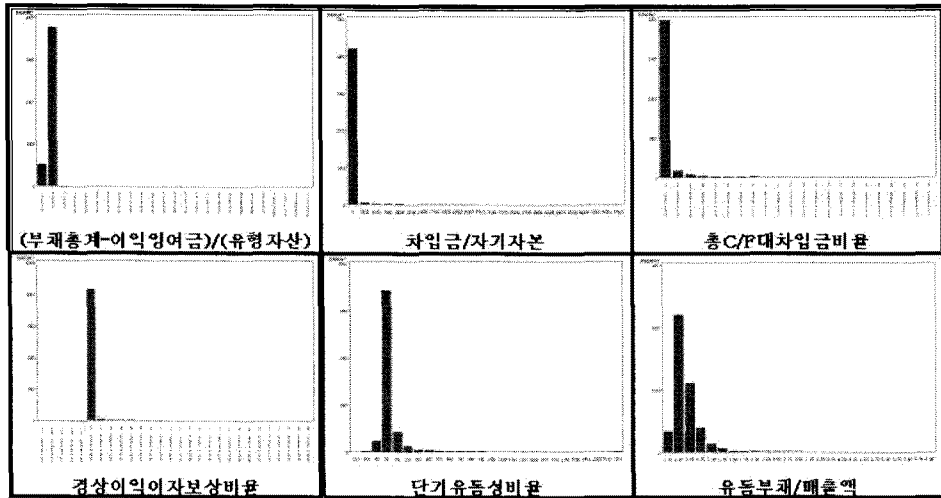


그림 3.1: 6개 변수의 히스토그램

재무비율에 대한 정확한 자료를 수집하는데 어려움이 있으나 범주형 형태의 자료수집이 가능한 경우에 모형을 개발할 수 있다. 또한 재무자료의 심한 왜도로 인해 발생할 수 있는 영향을 줄일 수 있다는 장점이 있다. 연속형 자료를 범주형으로 변환해 개발한 모형의 성과가 기존 신용평가모형의 성과와 비슷하다면, 범주형 자료는 정보손실이 적다고 판단되어 범주형 신용평가모형 자료로서 의미가 있을 것이다. 또한 Loess 방법 등 복잡하면서도 연도별 특수상황이 무시되고, 재무변수를 예측부도율로 대체하면서 발생하는 정보손실의 문제점도 보완할 수 있다.

### 3.2. 범주형 자료를 이용한 신용평가모형

그림 3.1에서 알 수 있듯이 많은 설명변수들에 왜도가 발생한다. 따라서 왜도의 발생을 줄이기 위해 연속형 자료를 여러 구간( $k = 3, 4, 5$  등)으로 나누어 범주화한다. 범주화의 기준은 연속형 자료의 왜도를 살펴본 후 구분하는데 자료의 특성상 등비율 방법을 이용하였으며 분위수를 통해 동일한 비율을 가지도록 등비율 구간(equal ratio interval)으로 범주화하였다. 3구간의 등비율로 나눌 때는 각 구간별 33.3%의 비율이 되도록 할당하였으며, 4구간일 때에는 각 구간이 25% 비율이 되도록 구간을 나누었다. 여러 구간으로 범주화한 자료들을 이용해 로지스틱 회귀분석을 실시해 보고 연속형 재무자료일 때의 신용평가모형과 비교한다.

#### 3.2.1. 3구간으로 범주화한 모형

연속형 변수를 앞에서 언급한 3구간의 등비율의 방법으로 범주화하였고, 3구간으로 범주화한 모형의 로지스틱 회귀분석 결과와 분류성적을 표 3.1에 나열하였다. 표 3.1을 살펴보면 최종변수에 선택된 회귀계수들이 모두 유의하지 않다는 것을 알 수 있으며, 3등비

표 3.1: 3구간으로 범주화한 신용평가모형 결과와 분류성과

변수		Estimate	Wald Chi-Square	p-값		
상수항		-4.6018	461.4459	<.0001		
건설업		-0.1692	0.5465	0.4597		
서비스업		-0.7087	2.2136	0.1368		
IT		0.5777	5.0472	0.0247		
(부채총계-이익잉여금) /(유형자산)	1	-0.2538	0.8317	0.3618	검증 통계량	분류성과
	2	-0.4080	5.2560	0.0219		
차입금/자기자본	1	-0.3054	0.4855	0.4859	1-PH	0.9481
	2	-0.2637	1.1055	0.2931	<i>K-S</i>	0.6717
총C/F대차입금비율	1	0.3511	1.7499	0.1859	AR	0.7677
	2	-0.5283	3.9281	0.0475	AUROC	0.8838
경상이익이자보상 비율	1	0.7218	8.7099	0.0032	Concordant	0.8630
	2	-0.4846	3.6980	0.0545		
단기유동성비율	1	0.7149	13.2244	0.0003		
	2	-0.1540	0.5635	0.4528		
유동부채/매출액	1	-0.5945	5.4979	0.0190		
	2	0.0806	0.2085	0.6479		

표 3.2: 4구간으로 범주화한 신용평가모형 결과와 분류성과

변수		Estimate	Wald Chi-Square	p-값				
상수항		-4.6387	436.4305	<.0001				
건설업		-0.4442	3.2519	0.0713				
서비스업		-0.8074	2.8118	0.0936				
IT		0.5812	4.5247	0.0334				
(부채총계-이익잉여금) /(유형자산)	1	-0.0764	0.0316	0.8590	검증 통계량	분류성과		
	2	-0.7564	8.4891	0.0036			<i>D</i>	1.4252
	3	-0.0263	0.0162	0.8987			1-PH	0.9654
차입금/자기자본	1	0.3986	0.4272	0.5134	<i>K-S</i>	0.6702		
	2	-0.4500	1.7680	0.1836	AR	0.7931		
	3	-0.4684	2.2322	0.1352	AUROC	0.8965		
총C/F대차입금비율	1	0.8183	6.8795	0.0087	Concordant	0.8860		
	2	-0.2484	0.5896	0.4426				
	3	-0.0709	0.0433	0.8351				
경상이익이자보상 비율	1	0.7247	7.3187	0.0068				
	2	-0.2909	1.0416	0.3074				
	3	0.1170	0.1410	0.7073				
단기유동성비율	1	0.7390	10.4111	0.0013				
	2	-0.1311	0.2812	0.5959				
	3	0.0788	0.0911	0.7628				
유동부채/매출액	1	-0.7982	5.4911	0.0191				
	2	-0.0490	0.0420	0.8376				
	3	0.3273	2.8174	0.0932				



표 3.3: 5구간으로 범주화한 신용평가모형 결과와 분류성과

변수		Estimate	Wald Chi-Square	p-값		
상수항		-4.6864	406.6518	<.0001		
건설업		-0.5965	4.8484	0.0277		
서비스업		-0.9250	3.5892	0.0582		
IT		0.5738	4.2613	0.0390		
(부채총계-이익잉여금)/(유형자산)	1	0.5084	0.9739	0.3237		
	2	-0.7532	4.6756	0.0306		
	3	-0.5267	4.8049	0.0284		
	4	-0.0132	0.0038	0.9511		
차입금/자기자본	1	0.2058	0.0714	0.7893		
	2	-0.7950	2.4467	0.1178		
	3	-0.2492	0.4539	0.5005	검증 통계량	분류성과
	4	0.0016	0.0000	0.9964	<i>D</i>	1.4284
총C/F대차입금비율	1	0.8389	5.9927	0.0144	1-PH	0.9676
	2	-0.2231	0.3915	0.5315	<i>K-S</i>	0.6839
	3	-0.1925	0.2486	0.6181	AR	0.7948
	4	0.0102	0.0006	0.9811	AUROC	0.8974
경상이익이자보상비율	1	0.8405	8.1386	0.0043	Concordant	0.8890
	2	0.0327	0.0120	0.9128		
	3	-40.6762	2.6685	0.1024		
	4	0.3008	0.6445	0.4221		
단기유동성비율	1	0.7607	9.0877	0.0026		
	2	0.0151	0.0031	0.9557		
	3	-0.0108	0.0013	0.9709		
	4	0.1614	0.2347	0.6280		
유동부채/매출액	1	-1.0913	6.3252	0.0119		
	2	-0.0578	0.0360	0.8495		
	3	0.1319	0.2651	0.6066		
	4	0.4206	3.9329	0.0474		

을로 범주화한 모형의 분류성과는 자유도에 의존하는 *D*와 *K-S*통계량뿐만 아니라 1-PH, AR, AUROC 그리고 Concordant 통계량들이 큰 값을 나타내기 때문에 모형의 예측률이 높다는 결론을 내릴 수 있겠다. 2절에서 토론한 기존의 신용평가모형과 여기에서 제안한 3구간으로 범주화한 재무자료의 신용평가모형의 비교는 4구간과 5구간으로 범주화한 모형의 결과와 같이 본절의 마지막에 비교 분석한다.

### 3.2.2. 4구간으로 범주화한 모형

연속형 변수를 4구간의 등비율 방법으로 범주화하고, 4구간으로 범주화한 자료의 로지스틱 회귀분석 결과와 분류성과를 표 3.2에 나열하였다. 3구간으로 범주화했을 경우와 마찬가지로 4구간으로 범주화한 신용평가모형의 로지스틱 회귀분석 결과를 보면 최종적으로 선택된 변수 모두가 유의하지 않고, 모든 분류성과 통계량들이 큰 값을 나타내기 때문에 모형의 예측률이 높다고 할 수 있다.

표 3.4: 기존 모형과 등비율 모형과의 분류성과 차이

통계량	기존 모형	등비율 모형과 차이					
		2구간	3구간	4구간	5구간	6구간	7구간
<i>D</i>	1.4035	-0.0740	-0.0273	0.0216	0.0249	0.0250	0.0439
1-PH	0.9722	-0.0297	-0.0241	-0.0068	-0.0046	0.0000	0.0032
<i>K-S</i>	0.6676	0.0054	0.0041	0.0026	0.0164	0.0110	0.0130
AR	0.7819	-0.0386	-0.0142	0.0112	0.0129	0.0129	0.0227
AUROC	0.8909	-0.0193	-0.0071	0.0056	0.0064	0.0065	0.0113
Concordant	0.8830	-0.0760	-0.0200	0.0030	0.0060	0.0060	0.0090

### 3.2.3. 5구간으로 범주화한 모형

연속형 변수를 5구간의 등비율의 방법으로 범주화하고, 5구간으로 범주화한 자료의 로지스틱 회귀분석 결과와 분류성적을 표 3.3에 나열하였다. 5구간으로 범주화한 신용평가 모형의 로지스틱 회귀분석 결과를 살펴보면 분류성적은 표 3.3에 열거한 모든 통계량들이 큰 값을 나타내기 때문에 모형의 예측력이 높다는 것을 파악할 수 있다.

재무비율을 예측부도율로 변환하여 사용한 신용평가모형(기존 모형)과 왜도가 심한 재무 자료를 범주화하여 개발한 신용평가모형(등비율 모형)의 분류성적을 비교해보자. 연속형 재무 자료를 3~5구간의 범주형 자료로 변환 후 본 연구에서 제안한 신용평가모형의 결과인 표 2.2, 3.1, 3.2 그리고 3.3 이외에 2구간과 6구간 그리고 7구간까지로 변환한 범주형 자료에 대해서도 모형을 개발한 결과를 포함하여, 기존 모형의 통계량 값과 등비율 모형의 통계량 값과의 차이를 표 3.4에 정리하였다.

표 3.4를 살펴보면 등비율 범주형 자료로 변환한 신용평가모형의 성과는 2절에서 토론한 연속형 재무자료에 대한 신용평가모형의 분류성과와 큰 차이를 보여주지 않는다는 것을 파악할 수 있다. 특히 4구간인 경우에 *D*, *K-S*, AR, AUROC 그리고 Concordant 통계량의 값이 연속형 재무자료에 대한 신용평가모형의 성과와 매우 유사한 값을 가졌으며, 그 외의 다른 검증 통계량인 1-PH 값도 4구간의 범주형으로 변환한 신용평가모형의 성과와 큰 차이가 없다. 그러므로 신용평가모형 개발에 연속한 재무자료를 범주화한 자료도 신뢰할 수 있으며, 범주화를 위한 가장 적합한 구간은 2~7구간 중에서도 4구간임을 파악할 수 있다.

## 4. 통계적 변수선택법을 이용한 신용평가모형

2장에서 설명한 기존의 신용평가모형 개발 과정에서 일변량 검정과 상관분석을 통해 변수선택을 한 후, 금융전문가들이 모형의 설명력, 선택된 변수의 이해가능성, 직관성, 모형의 안정성 등을 포괄적으로 고려하여 최종적으로 변수선택을 하면서 최종 신용평가모형을 구축하였다. 여기에서는 금융전문가의 의견을 수렴하지 못하고 오직 통계전문가들의 의견만으로 개발된 모형과 2절에서의 기존 모형을 비교하고자 한다.

2.3절 변수선택에서 선정된 12개의 변수들에 대하여 통계적인 변수선택 방법 중에서 단

표 4.1: 단계별 선택 모형 결과와 분류성과

변수	Estimate	Wald Chi-Square	p-값	검증 통계량	분류성과
상수항	-6.8184	462.0566	<.0001	<i>D</i>	1.4081
건설업	-0.5911	4.8038	0.0284	1-PH	0.9727
서비스업	-1.0296	4.3865	0.0362	<i>K-S</i>	0.6739
IT	0.5732	4.2489	0.0393	AR	0.7843
(자본총계/cpi)/ (금융비용/cpiavg)	0.1378	31.9778	<.0001	AUROC	0.8921
총C/F대차입금비율	0.0801	9.6142	0.0019	Concordant	0.8860
단기유동성비율	0.0901	14.3306	0.0002		
(부채총계/cpi-이익잉여금 /cpiavg)/(유형자산/cpi)	0.1660	15.7424	<.0001		
매출총이익/이자비용	0.0812	10.2922	0.0013		

계별 선택법(stepwise selection method)을 이용하여 변수들을 선택하고 로지스틱 회귀분석을 통해 새로운 신용평가 모형을 만들어 보고 적합성을 비교하였다.

#### 4.1. 단계별 선택법을 이용한 변수 선택

단계별 선택법을 이용하여 변수를 선택하였고 각 단계별로 변수를 투입하고 제거하기 위한 유의수준으로 0.05로 주었다. 최종모형에는 부채상환능력에 관련된 재무변수가 2개((자본총계/cpi)/(금융비용/cpiavg), 총C/F대차입금비율), 유동성에 관련된 변수가 1개(단기유동성비율), 안전성에 관련된 변수가 1개((부채총계/cpi-이익잉여금/cpiavg)/(유형자산/cpi)), 수익성에 관련된 변수가 1개(매출총이익/이자비용)로 총 5개의 재무변수가 선정되었다.

금융전문가들이 현업의 이해가능성 및 업계 경험상 필요한 변수들을 추가하거나 대체시켜서 얻은 기존 신용평가모형의 최종 6개 변수와 통계적 방법인 단계별 선택법에 의하여 선정된 5개 변수들을 비교해보면, 기존 신용평가모형의 변수들은 단계별 선택법에 의한 모형의 변수들과 중복되거나 연관성이 높은 변수들로 구성되어 있음을 알 수 있다. 3개의 가변수를 포함하면서 단계별 선택법에 의한 변수들의 신용평가모형의 결과와 분류성과는 표 4.1과 같이 나타난다.

표 4.1에서 얻은 분류성과 결과를 기존의 신용평가모형의 분류성과인 표 2.2와 비교해보면, 단계별 선택 모형의 모든 검증 통계량 값은 기존 모형의 값보다 약간 적은 값을 가지나 그 차이는 매우 작기 때문에 두 모형의 성과는 유사하며, 두 신용평가모형 모두 자료를 잘 설명하는 모형이라고 결론내릴 수 있다.

#### 4.2. 선택된 변수를 범주화한 신용평가모형

4.1절에서 단계별 선택법으로 선택된 변수들은 예측부도율로 변환한(2.2절 참조) 변수인데, 변환되기 전에 3.2절에서와 같이 대응하는 원자료의 재무변수를 3~5구간으로 범주

표 4.2: 3구간으로 범주화한 단계별 선택 모형 결과와 분류성과

변수		Estimate	Wald Chi-Square	p-값		
상수항		-4.5179	479.2466	<.0001		
건설업		-0.2554	1.2806	0.2578		
서비스업		-0.6973	2.1469	0.1429		
IT		0.6257	5.8833	0.0153	검증 통계량	분류성과
(자본총계/cpi)/ (금융비용/cpiavg)	1	0.5866	5.5137	0.0189	<i>D</i>	1.3654
	2	-0.4898	4.2381	0.0395	1-PH	0.9703
총C/F대출입금비율	1	0.5635	6.7221	0.0095	<i>K-S</i>	0.6678
	2	-0.5298	5.3179	0.0211	AR	0.7620
단기유동성비율	1	0.8837	22.2295	<.0001	AUROC	0.8810
	2	-0.1525	0.5653	0.4521	Concordant	0.8540
(부채총계/cpi-이익잉여금/cpiavg)/(유형자산/cpi)	1	-0.2850	0.9781	0.3227		
	2	-0.3839	4.5292	0.0333		
매출총이익/이자비용	1	0.7136	9.2356	0.0024		
	2	-0.1670	0.5647	0.4524		

표 4.3: 4구간으로 범주화한 단계별 선택 모형 결과와 분류성과

변수		Estimate	Wald Chi-Square	p-값		
상수항		-4.6220	429.7089	<.0001		
건설업		-0.5894	6.0048	0.0143		
서비스업		-0.8090	2.8165	0.0933		
IT		0.6118	5.1204	0.0236		
(자본총계/cpi)/(금융 비용/cpiavg)	1	0.6458	5.4896	0.0191	검증 통계량	분류성과
	2	-0.2270	0.6505	0.4199	<i>D</i>	1.4209
	3	-0.8194	4.8115	0.0283	1-PH	0.9659
총C/F대출입금비율	1	0.8639	11.5584	0.0007	<i>K-S</i>	0.6810
	2	-0.3055	1.2313	0.2671	AR	0.7909
	3	-0.1640	0.2873	0.5920	AUROC	0.8954
단기유동성비율	1	0.9624	20.5348	<.0001	Concordant	0.8790
	2	-0.0222	0.0086	0.9262		
	3	0.0848	0.1074	0.7432		
(부채총계/cpi-이익잉 여금/cpiavg)/(유형자 산/cpi)	1	-0.0822	0.0342	0.8534		
	2	-0.7637	8.4221	0.0037		
	3	-0.0198	0.0091	0.9241		
매출총이익/이자비용	1	0.8280	10.1218	0.0015		
	2	-0.2340	0.7121	0.3987		
	3	0.0634	0.0535	0.8171		

화하여 신용평가모형을 개발하고, 단계별 선택법을 이용한 신용평가모형과 비교해 본다.

우선, 4.1절에서 구한 신용평가모형에 포함된 연속형 변수들을 3구간의 등비율의 방법으로 범주화하고, 3구간으로 범주화한 모형의 로지스틱 회귀분석 결과와 분류성과를 표

표 4.4: 5구간으로 범주화한 단계별 선택 모형 결과와 분류성과

변수		Estimate	Wald Chi-Square	p-값		
상수항		-4.5914	428.7945	<.0001		
건설업		-0.6376	6.1151	0.0134		
서비스업		-0.7860	2.6204	0.1055		
IT		0.6480	5.5987	0.0180		
(자본총계/cpi)/(금융 비용/cpiavg)	1	1.0336	10.5909	0.0011		
	2	0.0479	0.0214	0.8837		
	3	-0.3853	1.1086	0.2924		
	4	-0.1568	0.1557	0.6931		
총C/F대출금비용	1	0.9675	13.1097	0.0003	검증 통계량	분류성과
	2	-0.2217	0.5898	0.4425	<i>D</i>	1.4311
	3	-0.3204	0.9616	0.3268	1-PH	0.9681
	4	-0.3053	0.5991	0.4389	<i>K-S</i>	0.6600
단기유동성비용	1	1.0500	20.7554	<.0001	AR	0.7961
	2	0.1515	0.3413	0.5591	AUROC	0.8981
	3	-0.0103	0.0012	0.9718	Concordant	0.8890
	4	-0.0170	0.0028	0.9582		
(부채총계/cpi-이익잉 여금/cpiavg)/(유형자 산/cpi)	1	0.3848	0.5603	0.4542		
	2	-0.7250	4.4135	0.0357		
	3	-0.4291	3.1908	0.0741		
	4	0.0504	0.0549	0.8147		
매출총이익/이자비용	1	0.6699	5.8840	0.0153		
	2	-0.0224	0.0061	0.9375		
	3	-0.4440	1.7111	0.1908		
	4	0.3348	1.0449	0.3067		

4.2에 나열하였다.

단계별 선택법의 변수들을 3구간으로 범주화한 신용평가모형의 결과에 따르면 3구간의 등비율로 범주화한 모형의 분류성과는 자유도에 의존하는 *D*와 *K-S* 통계량뿐만 아니라 1-PH, AR, AUROC 그리고 Concordant 통계량들이 큰 값을 나타내기 때문에 모형의 예측률이 높다는 결론을 내린다. 여기에서 제안한 단계별 선택법에 의하여 선정된 재무변수들을 3구간으로 범주화한 신용평가모형과 4.1절에서 토론한 단계별 선택법을 이용하여 설정된 신용평가모형과의 비교는 3구간 이외의 다른 종류의 구간으로 범주화한 신용평가모형들을 논의한 후 본절의 마지막에 토론한다.

단계별 선택법에 의해 선정된 연속형 변수를 4구간의 등비율 방법으로 범주화하고 4구간으로 범주화한 자료의 로지스틱 회귀분석 결과와 분류성과를 표 4.3에, 5구간으로 범주화한 자료에 대하여는 표 4.4에 나열하였다.

3구간으로 범주화했을 경우와 마찬가지로 4구간으로 범주화한 신용평가모형의 분류성과는 모두 표 4.3에 열거한 모든 통계량들이 큰 값을 나타내기 때문에 모형의 예측률이 높다고 판단할 수 있다. 또한 5구간으로 범주화한 신용평가모형의 로지스틱 회귀분석 결과를 살펴보면 분류성과는 표 4.4에 열거한 모든 통계량들이 큰 값을 나타내기 때문에 모형

표 4.5: 단계별 선택 모형과 등비율 모형의 분류성과 차이

통계량	단계별 선택 모형	등비율 모형		
		3구간	4구간	5구간
<i>D</i>	1.4081	0.0428	-0.0127	-0.0229
1-PH	0.9727	0.0024	0.0068	0.0046
<i>K-S</i>	0.6739	0.0061	-0.0070	0.0139
AR	0.7843	0.0222	-0.0066	-0.0119
AUROC	0.8921	0.0111	-0.0033	-0.0059
Concordant	0.8860	0.0320	0.0070	-0.0030

의 예측률이 높다는 것을 알 수 있다.

2절에서 토론했듯이 기존 신용평가모형은 일변량 검정, 상관분석 등 여러 단계를 통해 재무변수를 선택한 후에, 모형의 설명력, 직관성, 안정성 등을 고려하여 추가하거나 대체하여 최종적으로 변수들을 선택하였으나, 여기에서는 통계적 단계별 선택법을 이용해 새로운 변수들을 선정하였다. 그리고 새롭게 선택된 변수들에 대응하는 원자료의 연속형 재무 변수를 3~5구간의 범주형 자료로 변환한 모형들에 관한 분류성과인 표 4.2, 4.3 그리고 4.4을 통합하여 표 4.1의 결과와의 분류성과 차이를 표 4.5에 정리하였다.

표 4.5를 살펴보면, 등비율 신용평가모형의 성과는 단계별 선택법에 의한 신용평가모형의 분류성과와 큰 차이를 보여주지 않는다. 특히 4구간인 경우에 *D*, AR, AUROC 통계량의 값이 연속형 재무자료에 대한 신용평가모형의 성과와 매우 유사한 값을 가졌다. 그 외의 다른 검증 통계량 1-PH, *K-S*, Concordant 값도 4구간의 범주형으로 변환한 신용평가모형의 성과와 큰 차이가 없다. 그러므로 단계별 선택법으로 선정된 재무 변수들을 범주화한 자료도 신뢰할 수 있으며, 범주화를 위한 가장 적합한 구간은 3절에서 얻은 결과와 유사하게 3~5 구간 중에서도 4구간임을 판단할 수 있다.

## 5. 결론

신용평가모형은 경기 변동, 신용평가 절차의 변화, 자료의 질에 관한 문제 등 다양한 이유로 신용평가모형의 예측력 또는 안정성이 저해될 수 있으며, 특히 신용평가모형에 사용되는 연속형 재무자료는 최적 가중값 및 평균 부도율도 시간이 지남에 따라 변화할 수 있기 때문에 자료수집 및 최신 자료로 갱신하는 어려움 등의 한계성 그리고 재무자료의 특성상 심한 왜도의 발생이 생기는 문제점 등이 있다. 따라서 신용평가모형에 의해 예측된 부도율을 정기적으로 모니터링하고 그 적합성을 검증할 필요가 있다.

재무자료를 예측부도율로 변환하여 사용하는 기존의 신용평가모형의 문제점을 해결하는 방안으로 재무자료를 범주형으로 변환하여 신용평가모형을 개발하였다. 본 연구에서는 연속형 재무 변수를 다양한 범주형 변수로 변환하고, 이 변수들을 바탕으로 신용평가모형을 개발하여 기존의 신용평가모형에서 생기는 문제점을 보완하면서 개선된 모형등급 산출 방법을 제안하였다. 그리고 일반적인 신용평가모형은 변수선택 과정에 있어서 현업의 과거경험, 경제적 의미 해석 용이성 및 직관성, 모형의 안정성 등을 고려한 주관적 견해가 포

함된 변수선택 방법이 사용되었으나 본 논문에서는, 통계적 변수선택 방법을 적용하여 변수를 선정하여 신용평가모형을 제안하였다.

본 연구에서 제안한 신용평가모형들에 대하여 여러 가지 적합성 검증 방법들을 실행해 본 결과 등비율의 범주형 자료로 변환한 신용평가모형의 분류성과는 기존의 신용평가모형의 분류성결과와 큰 차이가 나타나지 않는다. 따라서 간단하게 범주화한 재무자료를 이용한 모형도 예측력 타당성이 좋으며, 특히 4구간으로 범주화한 자료에 대한 신용평가모형의 성과가 좋다고 판단할 수 있다. 단계별 선택법을 이용하여 구한 신용평가모형과 모형에 포함되는 재무 변수들을 범주화한 자료에 대한 신용평가모형을 비교해 보아도 역시 분류성결과에서 큰 차이를 보여주지 않으며, 특히 4구간인 경우에 적합성 검증 통계량의 값이 단계별 선택법에 의한 신용평가모형과 매우 유사한 값을 가졌다. 그러므로 재무 변수를 범주형 변수로 변환한 자료도 신용평가모형 설정에 있어서 의미있는 자료가 된다는 결론을 내릴 수 있으며, 특히 4구간의 범주형 자료로 변환하여 개발된 신용평가모형은 기존의 신용평가모형에 비교하여 분류성결과와 매우 유사하므로 4구간으로 범주화한 재무자료를 기반으로 제안한 신용평가모형이 신용평가 전문가에 의존하는 기존의 신용평가모형에 비하여 유사한 검증 효과가 있다고 판단할 수 있다. 또한 통계적 변수 선택방법을 적용한 신용평가모형을 범주형 자료로 변환한 신용평가모형과 비교하였을 때 분류성결과 차이를 보이지 않았고, 유사하게 4구간으로 범주화한 재무자료에 대한 신용평가모형이 가장 근사한 값을 보여주었다.

결과적으로 'By year method' 단계를 통해 Loess 방법을 이용한 변수변환 하는 과정에서 년도에 따른 재무비율의 연동흐름의 무시와 재무비율을 예측부도율로 대체하여 사용하는 등 복잡하고 정보손실을 가져오는 과정을 실행하지 않더라도, 단순하게 재무비율을 범주화한 재무자료를 사용함으로써 효율성을 높일 수 있다는 결론을 내릴 수 있다. 따라서 전문가가 아니라도 4구간의 등비율(예를 들어 상, 중상, 중하, 하)로 범주화하는 방법이 어렵지 않게 범주화 할 수 있으므로 본 연구를 통하여 4수준의 등비율로 범주화한 재무자료를 기반으로 신용평가모형을 개발하는 것을 제안한다.

## 참고문헌

- 검사지원국 (2004). <신용리스크 따라잡기>, 금융감독원.
- 박병수 (2004). 신용평가모형의 적합성 검증 방안 및 선결과제, <대구은행 경제리뷰>, 192, 52-65.
- 임종건 (2005). 신용평가시스템에 대한 적합성 검증 방법론 소개, <금융감독원, Risk Review>, 33-54.
- Cleveland, W. S. (1979). Robust locally-weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, 74, 829-836.
- Cohen, R. A. (1999). *An Introduction to PROC LOESS for Local Regression*, SAS Institute Inc., Cary, North Carolina, USA.
- Engelmann, B., Hayden, E. and Tasche, D. (2003). Measuring the discriminative power of rating systems, Discussion paper, *Series 2: Banking and Financial Supervision*, 1.

- Hand, D. J. and Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: A review, *Journal of the Royal Statistical Society*, **160**, 523–541.
- Joseph, M. P. (2005). A PD validation framework for basel II internal ratings-based systems, *Credit Scoring and Credit Control IX*.
- Stein, R. M. (2005). The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing, *Journal of Banking & Finance*, **29**, 1213–1236.
- Thomas, L. C., Edelman, D. B. and Crook, J. N. (2004). *Readings in Credit Scoring: Readings in Credit Scoring*, Oxford University Press, Oxford.

[2008년 6월 접수, 2008년 7월 채택]



# Validation Comparison of Credit Rating Models for Categorized Financial Data

Chong Sun Hong<sup>1)</sup>, Chang Hyuk Lee<sup>2)</sup>, Ji Hun Kim<sup>3)</sup>

## Abstract

Current credit evaluation models based on only financial data except non-financial data are used continuous data and produce credit scores for the ranking. In this work, some problems of the credit evaluation models based on transformed continuous financial data are discussed and we propose improved credit evaluation models based on categorized financial data. After analyzing and comparing goodness-of-fit tests of two models, the availability of the credit evaluation models for categorized financial data is explained.

*Keywords:* Credit rating model; categorical data; financial data; performance; stepwise variable selection; validation.

- 
- 1) Professor, Department of Statistics, Sungkyunkwan University, 3-53, Myungryun-dong 3, Jongro-gu, Seoul 110-745, Korea. Correspondence: cshong@skku.ac.kr
  - 2) Researcher, Research Institute of Applied Statistics, Sungkyunkwan University, 3-53, Myungryun-dong 3, Jongro-gu, Seoul 110-745, Korea.
  - 3) Researcher, Research Institute of Applied Statistics, Sungkyunkwan University, 3-53, Myungryun-dong 3, Jongro-gu, Seoul 110-745, Korea.