
Trillion (1조)



재무비율을 이용한 부도예측에 대한 연구

정기호 이주희 신문혁 윤영주



CONTENTS

I. 논문 소개

논문 개요 | 주요 개념 정의

II. 선행 연구

선행 연구 소개

III. 분석

데이터 전처리 | 변수 유의성 검정 | Stepwise | Modeling | Validation | 논문 결과 분석

IV. 고찰

기대효과 | 한계점/향후 연구방향

V. 향후 계획

적용방향

I. 논문 소개

01. 논문 개요

“ 재무 비율을 이용한 부도 예측에 대한 연구 ”

저자 : 박종원, 안성만

재무비율을 이용한 부도예측에 대한 연구:
한국의 외부감사대상기업을 대상으로*

목적 : 한국 시장의 외부감사대상기업을 대상으로 재무비율을 이용한 부도예측모형을 개발하고 모형의 정확도와 변별력을 검증

박종원(교신저자)
서울시립대학교 경영대학 교수
(parkjw@uos.ac.kr)
안성만(주저자)
농협은행 여신정책부
(asm1230@nonghyup.com)

Keyword : 외부감사대상기업, 부도예측, 재무비율, 다변량 로짓분석

본 연구는 한국 시장의 외부감사대상기업의 부도예측을 위한 재무비율을 이용한 부도예측모형을 개발하고 모형의 정확도와 변별력을 검증하였다. 상장기업들을 대상으로 이루어진 선행연구와 달리 외부감사대상기업 전체를 대상으로 재무비율을 이용한 부도예측 모형을 구축하고 그 적용가능성을 살펴본다는 점에서, 기업의 부실을 설명해줄 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.

I. 논문 소개

02. 관련 개념 설명

외부 감사 대상 기업 선정

- 외부감사를 통해 신뢰성이 검증된 재무제표를 얻을 수 있는 외부감사대상기업으로 연구 대상 선정

- 외부감사 기업 대상 재무비율을 이용한 부도예측모형은 모형 구축의 용이성과 비용 측면에서 재무제표정보가 갖는 정보의 예측력과 관련하여 실무적으로나 학문적으로 중요한 연구대상이다.

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 설명해줄 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.

I. 논문 소개

02. 관련 개념 설명

부도 기준 2005년 1월 1일부터 2007년 12월 31일까지의 금융결제원 당좌거래정보를 이용하여
당좌수표정지 또는 약속어음부도로 인해 당좌거래가 정지된 기업

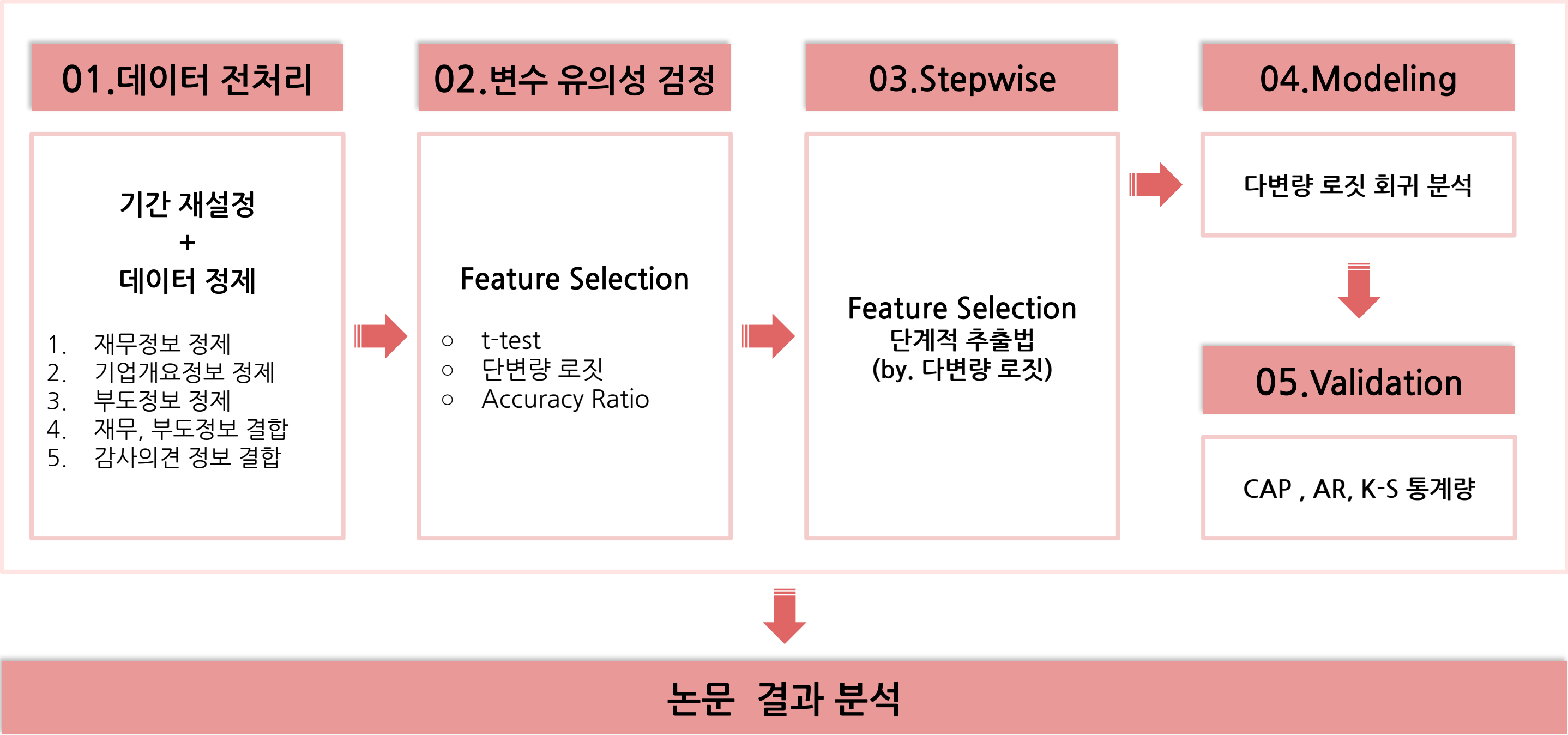
	2001년	2002년	2003년	2004년	2005년
			<div>⋮ ↓ 부도</div>	<div>⋮ ↓ 부도</div>	<div>⋮ ↓ 부도</div>
방법 1	정상	정상	부도	정상	부도
방법 2	정상	정상	부도	X	X
방법 3	정상	정상	부도	부도	부도

II. 선행 연구

연구자	연도	표본 기업	연구방법	연구결과
Beaber	1996	상장기업 (부도율 50%) 부도 79 정상 79	단일변량분석	이원분류 검정을 통한 분석결과 현금흐름/ 총부채비율의 예측력은 부도 1년, 5년 전에 각각 87, 78%로 가장 좋은 예측 지표로 밝혀졌으며, 그 다음으로는 총자산순이익률의 예측력이 가장 높게 나타남%
Ohlson	1980	상장 제조업 (부도율 4.85%) 부도 105 정상 2058	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 (부도율 50%) 부도 29 정상 29	다변량 판별분석	31개 재무비율과 3개의 실수 중 통계적 유의성이 있는 지표로 매년 4개의 주성분을 판별함수로 도출한 결과 부도 1년, 2년, 3년 전에 각가 94.83%, 77.59%, 75.86%의 예측력을 보임
이계원	1993	상장 기업 (부도율 22%) 부도 34 정상 123	로짓분석	ROE변수만을 사용한 기본모형과 여기에 변수를 추가한 확장모형으로 연구한 결과 기본모형의 예측력은 부도 1년 전의 경우 예측 정확성은 80%였으나 확장모형은 82%로 다소 높게 나타남.
Altman	1996	상장 제조/도소매 (부도율 36%) 부도 34 정상 61	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 높게 나타남
장휘용	1998	상장기업 (부도율 25%) 부도 34 정상 103	로짓분석	설명변수 7개로 모형을 구축하였는데, 자체예측력은 부실표본 에서 85.3%, 정상표본에서 5.1 %였으며 검정용표본에서는 부실과 정상표본에서 각각 76.5%와 94.2%의 예측력을 보임.
박정윤	2000	상장기업 (부도율 26%) 부도 41 정상 118	로짓분석	기업의 주요 재무정책변수 중 부실 4~5년 전에는 자기자본순 이익률과 총자본부채비율이 부실에 영향을 주나, 부실 전 3년 간은 현금흐름과 배당률의 크기가 부실에 영향을 주는 것으로 나타남.

III. 분석

Analysis Flow

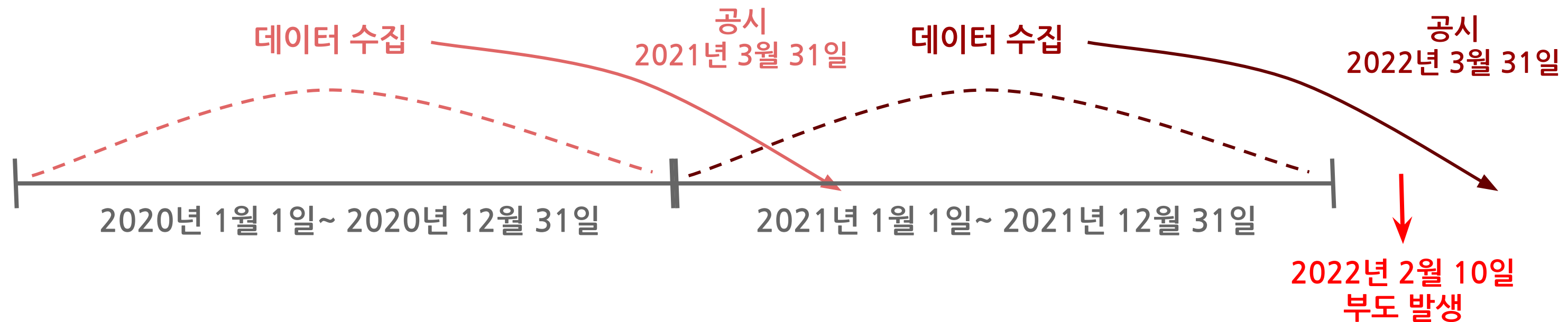


III. 분석

01. 데이터 전처리

기간 재설정

금융기관과 공공기관을 제외한 기업 중 공인회계사의 외부감사 의견이 있는 12월 말 결산법인인 17,913개 기업의 2001년 1월 1일부터 2006년 12월 31일까지 결산 재무제표를 이용했다



결산 재무제표가 공시되는 시점을 고려하지 않을 경우,
부도의 원인이 되는 사건이 발생한 기간의 결산자료를 이용해 부도를 예측하는 오류가 발생

➡ 재무제표가 공시되는 시점을 고려하여 당해년도 4월부터 이듬해 3월까지 12개월을 부도 관측기간으로 정의

III. 분석

01. 데이터 전처리

Step1. 재무정보 정제

재무정보 raw_data
905,480건

- 결산 일자 변경 등에 의해 회계기간이 1년 미만인 재무정보 삭제
- 연속 2개년 재무제표가 존재하지 않는 경우 삭제
- 결산 월이 12월이 아닌 데이터 삭제

정제 된 재무정보
875,663건

재무비율 파생 변수 생성

Step2. 기업개요정보 정제

기업정보 raw_data
835,666건

- 산업분류코드가 잘못 기입된 경우 데이터를 삭제
- 법인번호 또는 사업자번호가 중복된 개요정보 삭제

정제된 기업정보
833,190건

결합

기업정보가 없는 재무비율 데이터 삭제

data
865,420건

III. 분석

01. 데이터 전처리

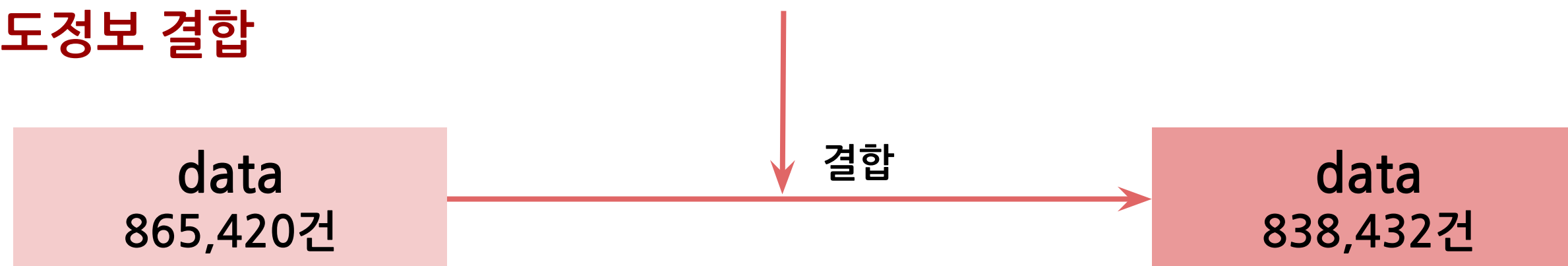
Step3. 부도정보 정제

부도정보 raw_data

- 전체 부도 차주는 개인과 법인을 모두 포함, 그 중 법인만 선별
- 최초 부도 시점을 부도로 인식하고, 그 이후 부도 데이터는 모집단에서 삭제

정제된 부도정보
3534건

Step4. 재무, 부도정보 결합



III. 분석

01. 데이터 전처리

Step5. 감사의견 정보 결합

재무정보와 부도정보 결합 후 data

- 회사 설립일 이전의 재무정보 삭제
- 감사의견 결합
- 제조/ 건설/ 도소매/ 서비스업에 해당하지 않는 기업 삭제
- 2003~ 2006년 12월 결산 법인만 선택
- 비외감 기업 삭제

최종 분석 모집단

- 부도기업 322개 포함 총 52,052개
- 주요 변수는 총 117개 재무비율

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	11,432	34	11,466	22%	0.30%
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	12,496	109	12,605	24%	0.86%
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	13,417	100	13,517	26%	0.74%
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	14,385	79	14,464	28%	0.55%
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	51,730	322	52,052	100%	0.62%

III. 분석

02. 변수 유의성 검정

Feature Selection

변수 유의성 검정

통계적 추정값의 신뢰도를 확인하기 위하여 통계적 이론에 근거하여 추론하는 통계검정

- ▶ T-검증

모집단의 분산과 표준편차를 알지 못할때,
모집단을 대표하는 표본으로부터 추정된 분산이나 표준편차를 가지고 검정하는 방법

- ▶ 단변량 로짓

로지스틱 회귀분석이란 종속변수가 명목형인 경우 적용되는 모형
직접적인 수치를 예측하는 선형회귀분석과는 달리 목표변수 Y가 특정 범주에 속할 확률을 구해줌.
이 중 단변량 로지스틱은 명목형인 종속변수가 1개인 로지스틱 회귀를 말함.

- ▶ Accuracy Ratio

측정된 값이 참값과 얼마나 일치하는지 나타내는 척도

III. 분석

02. 변수 유의성 검정

Feature Selection

변수 유의성 검정 [t-test]

정규 분포를 따르는 모집단의 모평균을 알고, 모분산을 알지 못하는 경우 사용하게 된다.
T-test를 수행하는 경우는 **두 모집단 간의 feature 평균 차이가 통계적으로 유의한지 검정**하게 되는 것이다.

소표본 (표본의 크기가 30개 미만)	T- test 사용 대신 모집단의 정규분포성, 모분산의 동분산성을 만족한다는 것을 가정
대표본 (표본의 크기가 30개 이상)	Z- test 사용

- 귀무가설 (H0) : 두 모집단 평균 차이가 유의하지 않다.
- 대립가설 (H1)
- ▶ 양측검정 : 두 모집단 평균 차이가 유의하다. (유의수준이 5%인 경우, 단방향에서 2.5%의 통계치로 검정한다)
 - ▶ 단측검정 : 어느 한 모집단의 평균이 크거나 작다. (유의수준이 5%인 경우, 그대로 5%로 검정한다)



해당 논문에서는 t-test를 활용하여 부도기업 집단(모집단1), 정상기업 집단(모집단2) 내에서 한 feature를 선택한 후, 분포를 확인하여 두 집단 간 재무비율의 차이가 통계적으로 유의미한지 검정하게 된다.
이 과정을 따라서 P-value < 0.05로 도출된 재무비율을 선택하게 되는 것이다.

III. 분석

02. 변수 유의성 검정

Feature Selection

변수 유의성 검정 [단변량 로짓]

feature(설명변수)의 개수	single(단순)	feature 1개 (ex. 단순선형회귀모형)
	multiple(다중)	feature 2개 이상 (ex. 다중선형회귀모형)
label(종속변수)의 개수	univariate(단변량)	label 변수가 1개 (단변량 로짓분석)
	bivariate(이변량)	label 변수가 2개 (이변량 로짓분석, T or F)
	multivariate(다변량)	label 변수가 2개 이상 (다변량 로짓분석, 신용등급 분류)



해당 논문에서는 단변량 로짓을 이용하여 각 feature들이 종속변수에 유의한 의미를 가지는지 분포를 이용하여 확인하며, 카이제곱 검정을 이용하여 모형의 적합성을 검정한다.
따라서 해당 논문에서는 t-test로 선별한 feature를 대상으로 단변량 로짓 분석을 시행하여 다시 한 번 feature를 필터링 하였다.

III. 분석

03. Stepwise

Feature Selection

단계적추출법 (Stepwise)

- ▶ 전진선택법 (Foreward Selection)

기존 모형에 가장 설명력이 좋은 변수를 하나씩 추가하는 방법

- ▶ 후진소거법 (Backward Elimination)

모든 변수가 포함된 모형에서 설명력이 가장 작은 변수를 제거해나가는 방법

- ▶ 단계적 선택법 (Stepwise Selection) : 전진선택법 + 후진 소거법

전진선택법에 의해 변수를 추가하고 새롭게 추가된 변수를 포함하여 모형에 적합합니다.

그리고 추정된 회귀 변수에 대하여 가장 큰 p-value 값을 구합니다.

최대 p-value값이 사전에 정의된 유의 수준보다 크거나 같으면 해당 변수를 제외하고 전진선택법을 진행하고

그렇지 않은 경우 제외하는 변수 없이 전진선택법을 진행합니다.

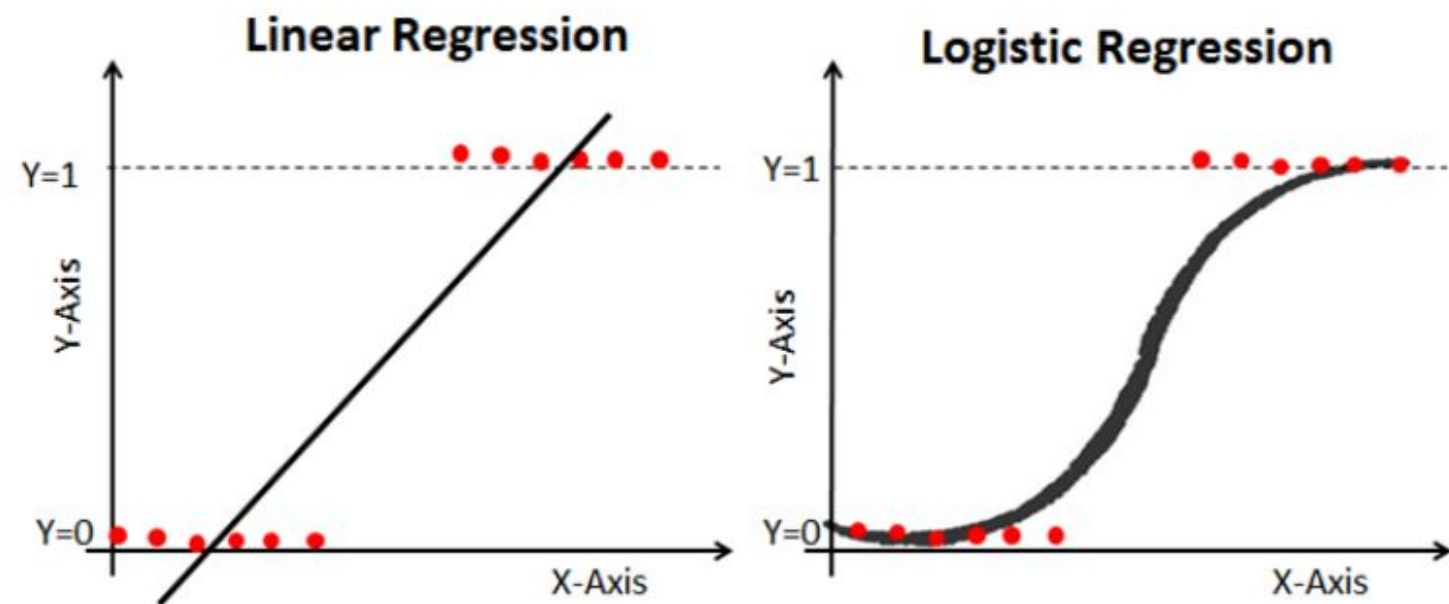
III. 분석

04. Modeling

다변량 로짓 회귀 분석

로지스틱 회귀 분석

분석하고자 하는 대상이 두집단 혹은 그 이상으로 나누어진 경우, 개별 관측치들이 어느 집단으로 분류될 수 있을지를 판단하는 분석방법.
독립변수는 연속형, 범주형 자료로 분석이 가능하며, 종속변수는 오직 범주형 자료로만 분석이 가능.



Feature의 크기와 무관하게 label과 선형결합을 이루고 있음을 가정한다.
label이 0,1 사이의 값으로 도출된다는 보장을 하지 못한다
Feature를 scaler해도 선형회귀의 기본가정인 오차항의 정규분포성 가정 충족하지 못함.
오차항의 이분산성 존재로 선형회귀의 동분산성 가정에 어긋난다.



부도예측에 사용될 feature는 재무비율 뿐 아니라 총자산 항목도 포함된다.
기업규모와 부도확률간의 관련성을 통제하기 위해 **총자산을 log 변환**한다.
이는 **기업 규모에 따른 오분류를 낮출 수 있지만, 선형회귀에서 사용하게 되면 오차항의 정규성 가정에 어긋나게 된다.**
따라서 해당 논문은 부도 예측 모형에 로짓 분석을 사용하게 되었다.

III. 분석

04. Modeling

다변량 로짓 회귀 분석

로지스틱 회귀 분석 과정

가설 설정 → 종속변수 범주화 → 모형 적합도 확인(χ^2) → 로지스틱 가설 검정

▶ 가설 설정

- H_0 : 독립변수가 종속변수에 영향을 미치지 않는다
- H_1 : 독립변수가 종속변수에 영향을 미친다

▶ 모형 적합도 확인

카이제곱 분석 결과인 p-value 확인으로 판단
 $p\text{-value} \leq 0.05$: 모델이 유의미

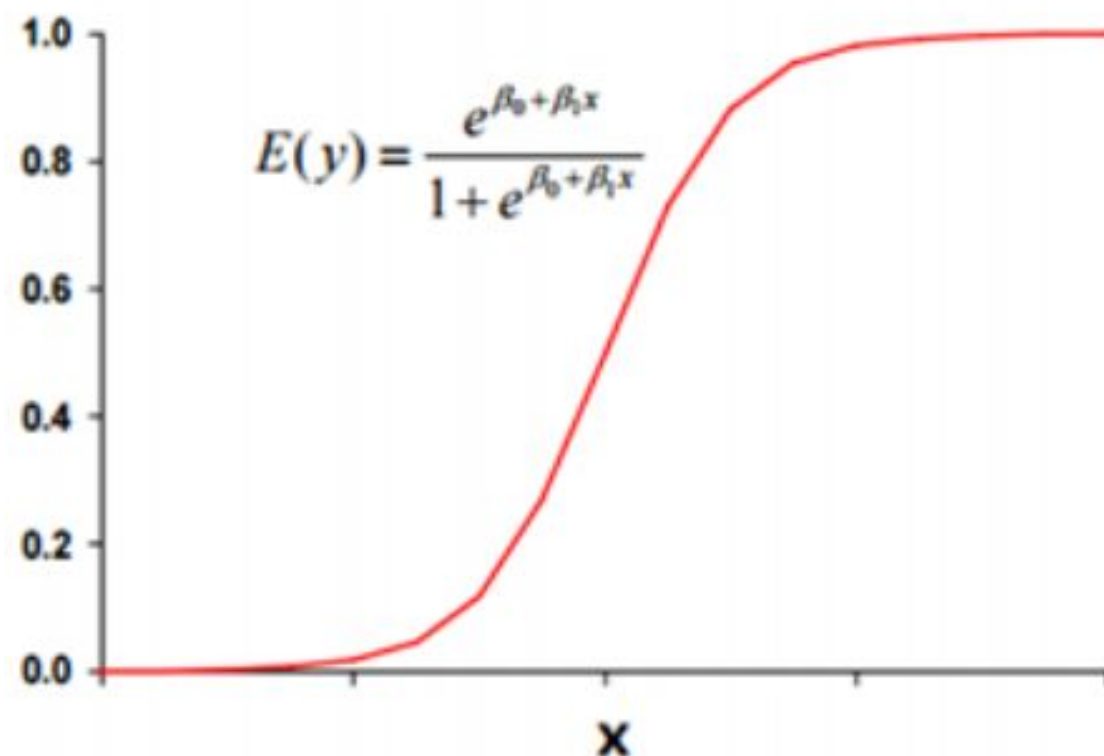
III. 분석

04. Modeling

다변량 로짓 회귀 분석

로지스틱 회귀 모형

로지스틱 회귀 모형이란, 독립변수가 종속변수 혹은 결과값에 상관 없이 항상 범위 [0,1] 사이에 있도록 하는 모형. 먼저 오즈비를 만들어주고 여기에 로짓변환을 시켜줌으로써 로지스틱 회귀모형을 생성.



우도 (Odds) : 일어날 확률 / 일어나지 않을 확률

$$Odds\ ratio = \frac{P_i}{1 - P_i}$$

로짓 변환 : Odds에 로그를 취한 형태

$$\ln\left(\frac{P_i}{1 - P_i}\right) = \beta X_i + u_i$$

Logistic Function

$$LF = \frac{\exp(\beta X_i + u_i)}{1 + \exp(\beta X_i + u_i)}$$

로지스틱 회귀 모형의 적합도는 우도비 검정을 이용.

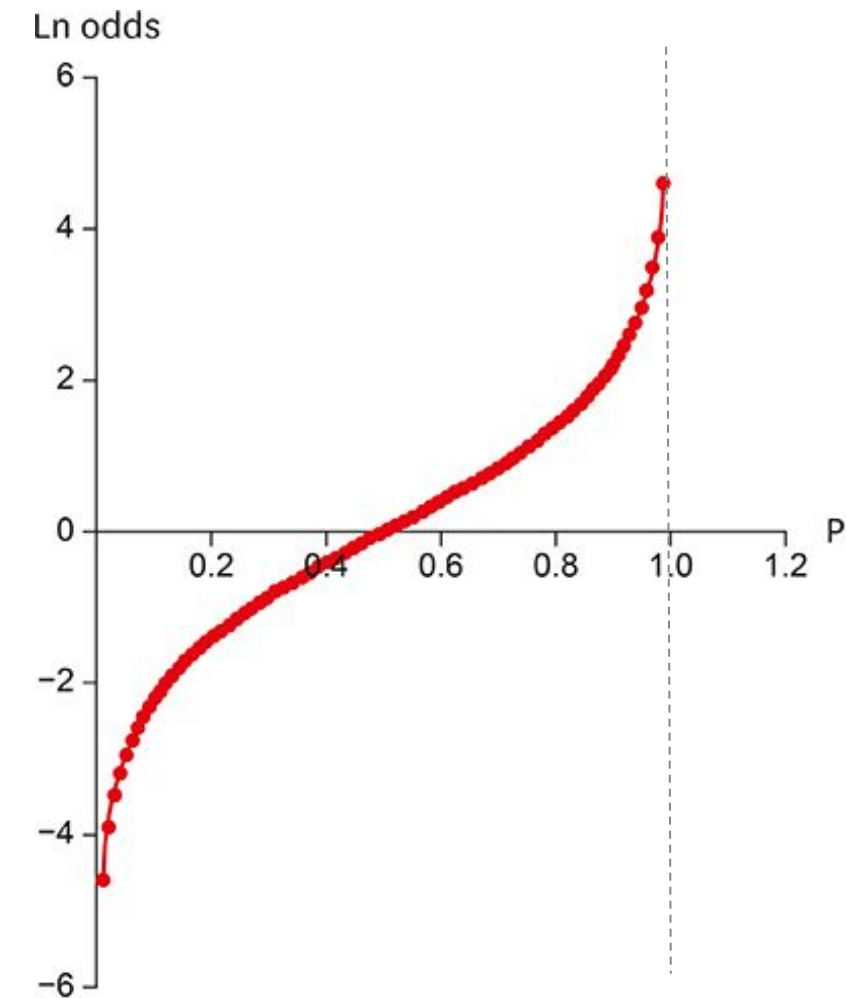
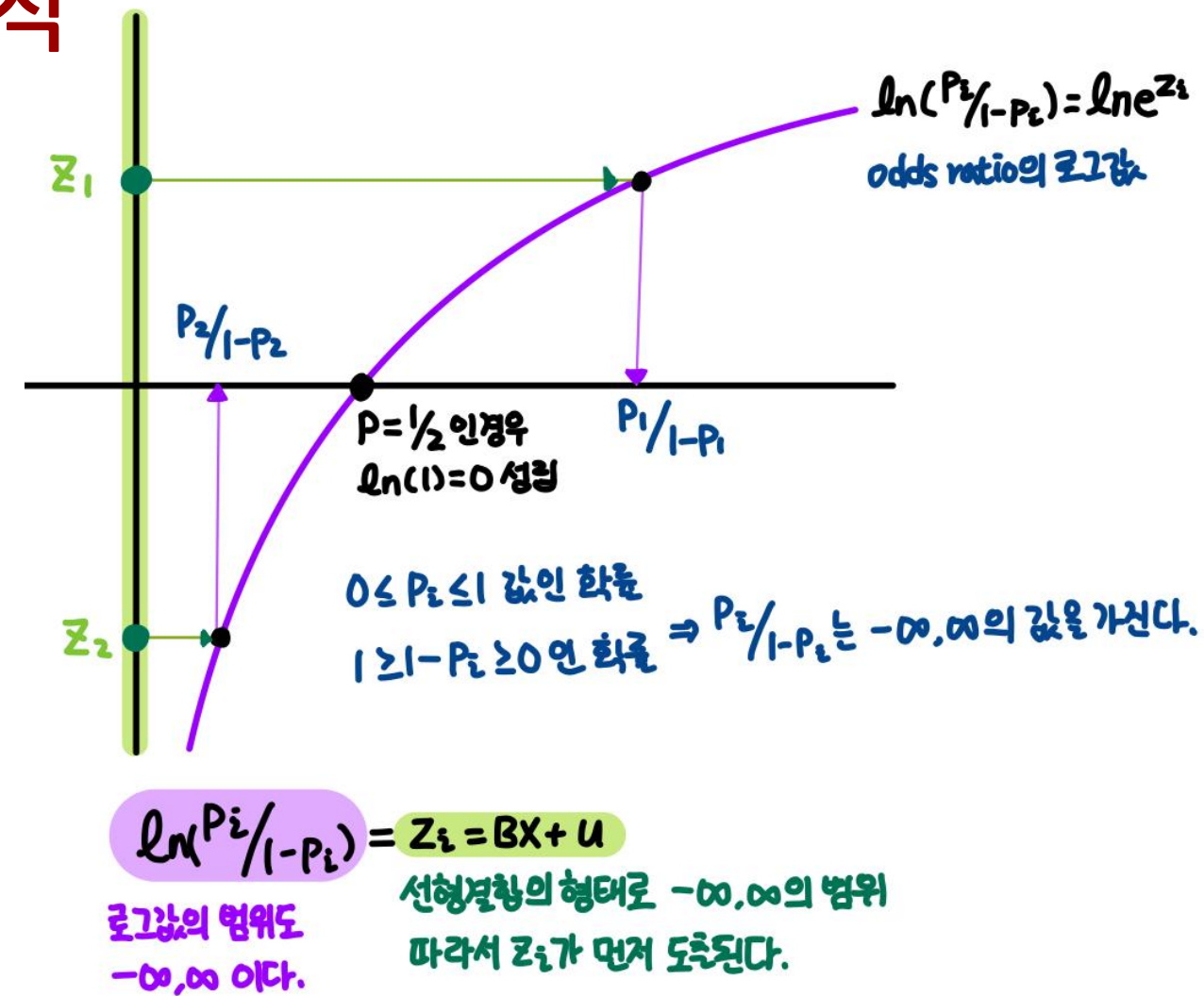
최대 우도법이란, 주어진 독립변수들로부터 종속변수를 가장 잘 예측하는, 즉 우도가 가장 높은 회귀식을 추정하는 방법.

III. 분석

04. Modeling

다변량 로짓 회귀 분석

로짓 모형



- P 는 $(0,1)$ 사이 값을 가지므로, feature인 X 가 결합된 **logit**은 $(-\infty, +\infty)$ 범위를 가질 수 있다. 따라서 logit 값이 무한정이다.
- **$L = BX + u$ 로 선형관계**이지만, **P 는 X 와 비선형관계**이다.
- $P/(1-P)$ 는 $P = 1/2$ 일 때, 1이 된다. 이때, $L = 0$ 으로 된다. 해석하자면, L 이 양의 값으로 증가할 때, 사건이 발생할 가능성은 높아진다.
- Odds ratio 로그 값이 L 이므로 **$B(X$ 의 계수)의 의미**는 feature가 한단위 변화할 때, **Odds Ratio의 로그값이 변화하는 정도**를 의미한다.
- $L = BX + u$ 에서 B 를 추정하게 되면 Odds ratio를 활용해 특정 기업이 부도일 확률을 구할 수 있다.

III. 분석

04. Modeling

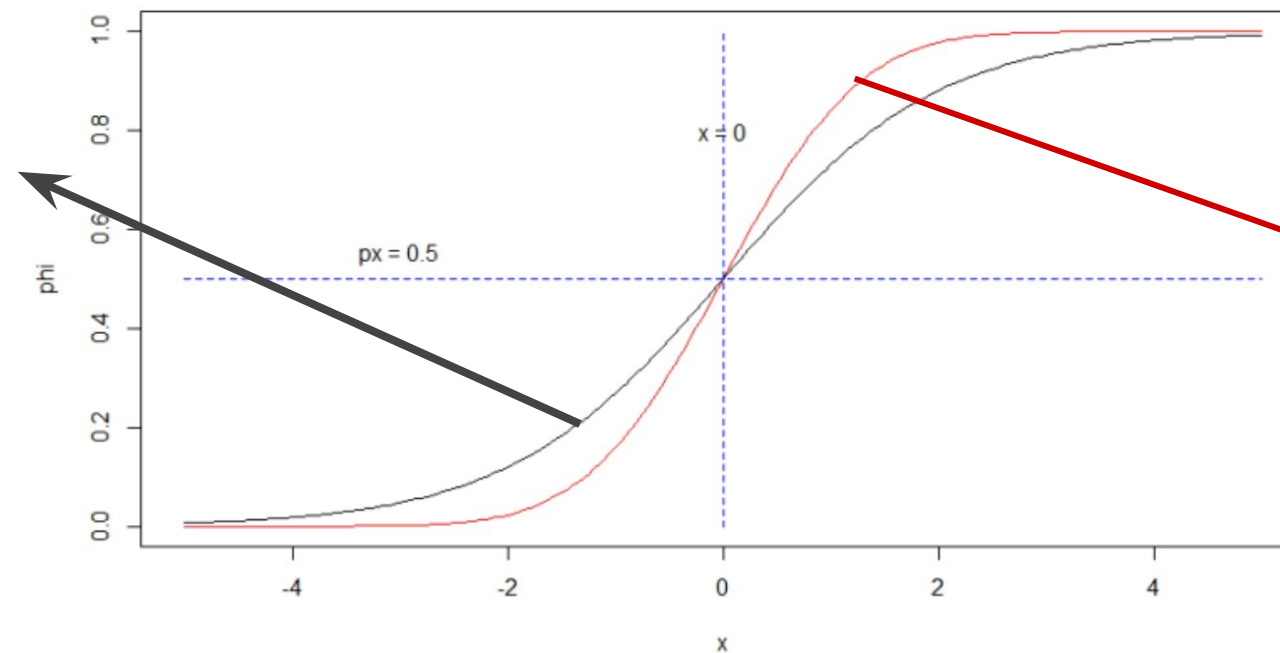
별첨

로짓, 프로빗분석

로짓 분석

- 오차항이 **로짓분포**를 따름
- 표준화된 로짓분포의 분산 값

$$Var(LF) = \frac{\pi^2}{3}$$



프로빗 분석

- 오차항이 **정규분포**를 따름
- 표준화된 정규분포의 분산 값

$$Var(probit) = 1^2$$

▶ 차이점

- 오차항에 대한 가정이 차이로 두 모형의 직접 비교는 변수의 추정치에 전환율을 곱해줘야 한다. (전환율 : 약 1.81 = $Var(LF)$ 의 루트값)
- **로짓모형의 분산이 더 크기 때문에 함수의 끝 부분이 더 두툼하다.** (프로빗이 더 얇아서 0,1에 더 빠르게 수렴한다.)

▶ 결론

추정하는 방식은 유사하게 진행되어 모형을 해석하는 부분에서 큰 차이는 없다.

실증분석에서는 프로빗과 로짓의 큰 차이는 없으나, **사회과학분야에서 분석을 진행할 때 로짓모형이 범용적으로 사용된다고 한다.**

III. 분석

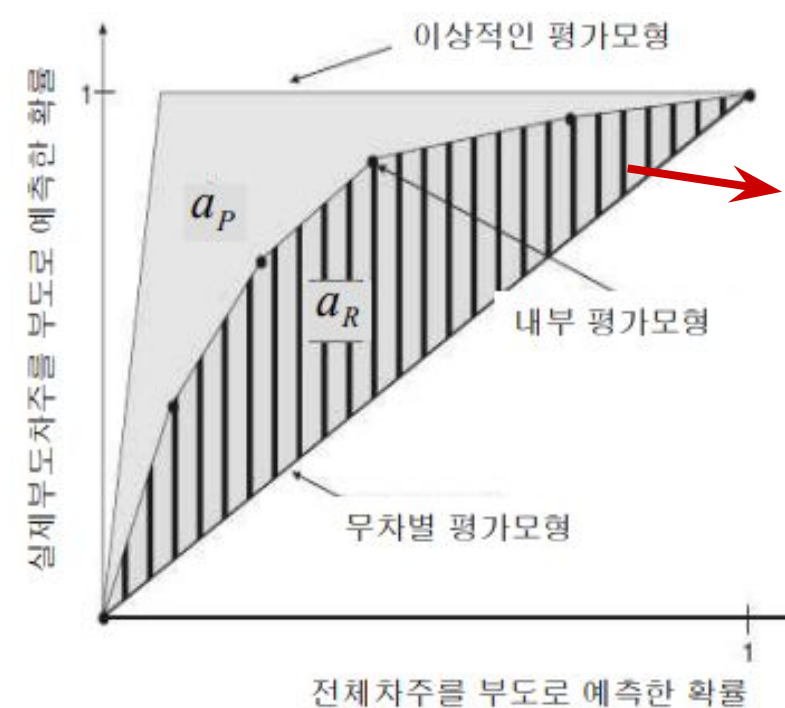
05. Validation

CAP, ROC

모형의 변별력 평가방법

▶ CAP(Cumulative Accuracy Profile) / AR(Accuracy Ratio)

X 축: 전체 기업 중 부도로 예측된 기업이 차지 하는 비중
Y축: 실제 부도기업 중 부도로 예측된 기업이 차지하는 비중

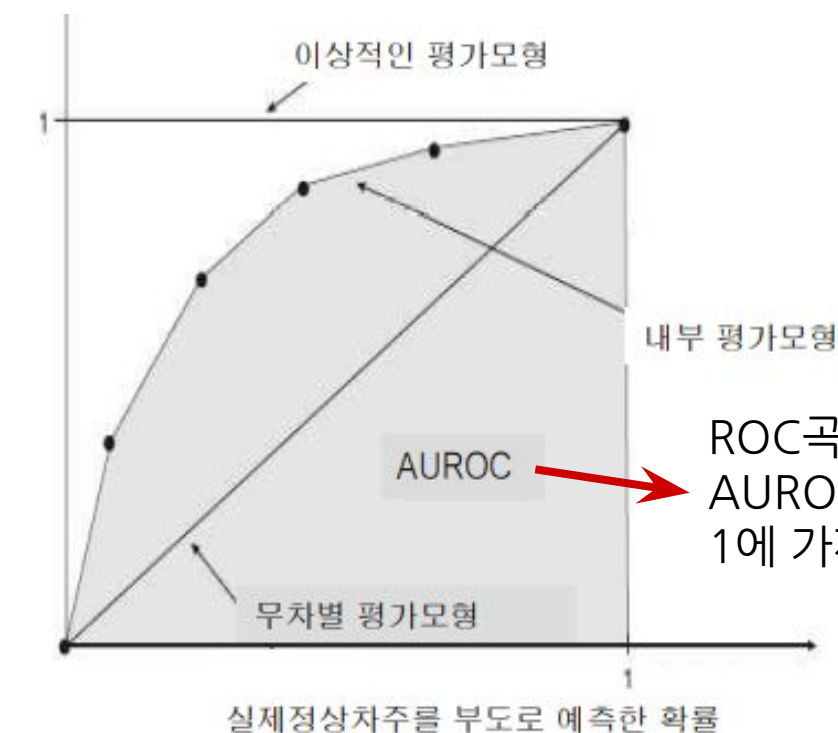


$$AR = \frac{a_R}{a_P}$$

AR값은 0과 1사이에 존재하며,
1에 가까울수록 변별력이 높음

▶ ROC(Receiver Operating Characteristics) / AUROC(Area Under ROC)

X 축: 전체 기업 중 부도로 예측된 기업이 차지 하는 비중
Y축: 실제 정상인 기업 중 부도로 예측된 기업이 차지하는 비중



ROC곡선 아랫 부분의 면적
AUROC값은 0.5에서 1사이의 값을 갖는다.
1에 가까울수록 변별력이 높음을 의미한다.

III. 분석

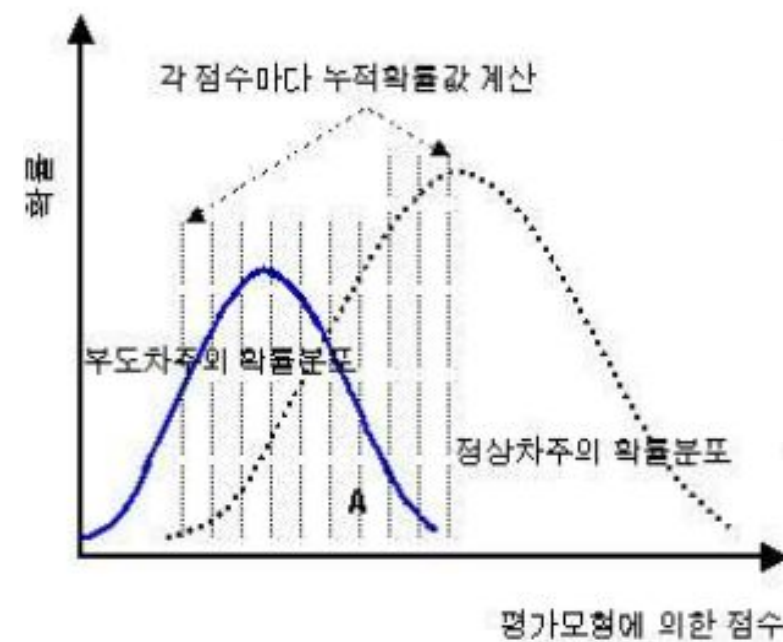
05. Validation

K-S 통계량

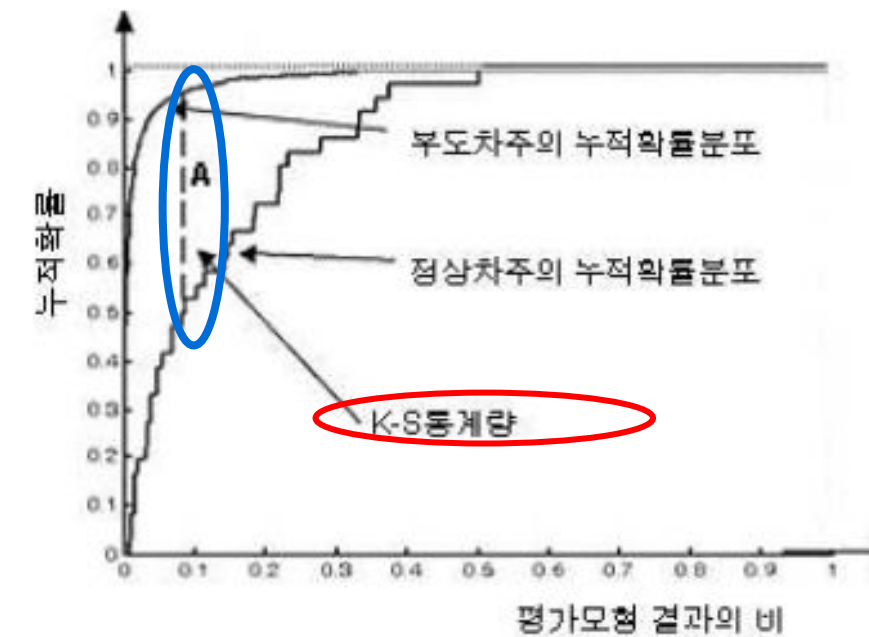
모형의 변별력 평가방법

▶ Kolmogorov-Smirnov(K-S) 통계량

정상기업과 부도기업을 구분하여 누적확률분포를 계산



각 등급별로 누적확률분포 차이를 계산



K-S통계량: 부도차주의 누적확률분포 - 정상차주의 누적확률분포

가장 큰 K-S 통계량을 가지는 값 선택

III. 분석

06. 논문 결과 분석

단일변량분석 결과 (t-test, 단변량 로짓모형, AR)



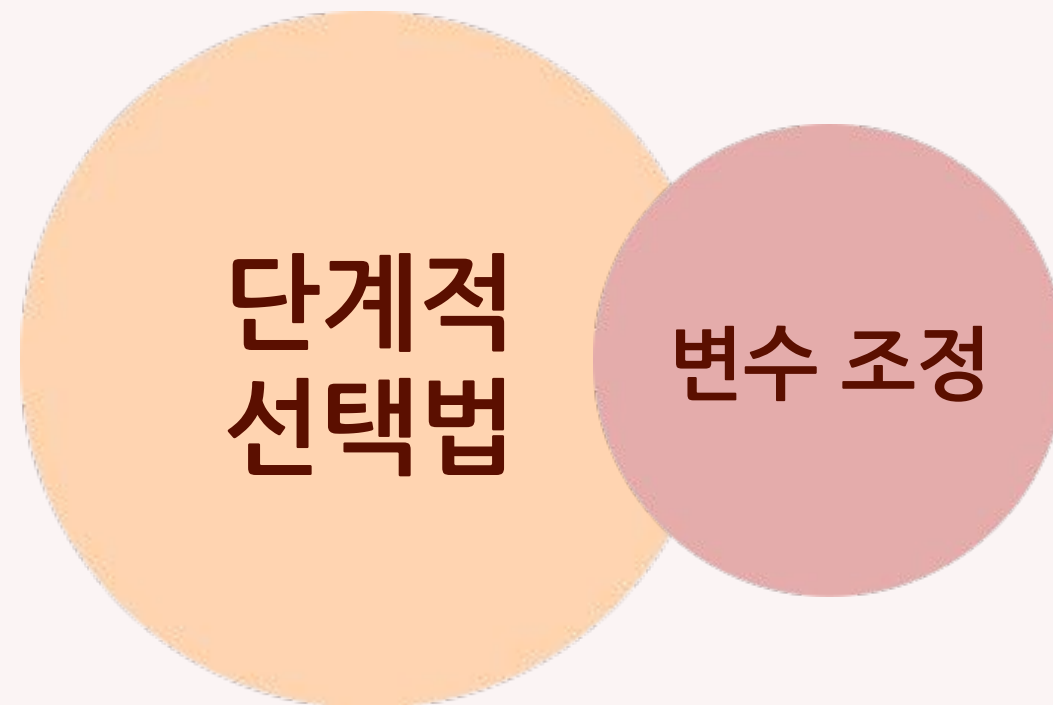
과도한 이자비용 부담이나
유동성관리에 실패하는 경우
부도위험에 노출

기업의 성장성이나 수익성 요인보다는,
안정성과 유동성 요인이
기업부실에 보다 중요한 영향을 끼침을 알 수 있음

III. 분석

06. 논문 결과 분석

다변량분석을 위한 전처리



- 1.업종별 부도확률의 차이를 감안하여 업종을 더미(dummy)변수로 처리
- 2.기업규모와 부도확률간의 관련성을 통제하기 위해 총자산에 자연로그(ln)를 취한 값을 변수로 투입
- 3.재무비율 자료에 존재하는 극단치와 결측치를 조정
 - > 결측치 : 해당 변수 중앙값으로 대체
 - > 극단치 : 평균에서 $-2 \times \text{표준편차}$ 를 벗어나는 경우 $\text{평균} - 2 \times \sigma$ 값으로, 평균에서 $+2 \times \text{표준편차}$ 를 벗어나는 경우 $\text{평균} + 2 \times \sigma$ 값으로 조정

III. 분석

06. 논문 결과 분석

다변량분석결과

로짓회귀식

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	자본금회전을	-0.0159	0.0059	7.1260	0.0076
FR090	안정성	차입금의존도	-0.0169	0.0036	21.809	< 0.0001
FR092	활동성	총부채회전을	-1.4934	0.2303	42.037	< 0.0001
FR095	생산성	총자본투자효율	-0.0561	0.0089	39.837	< 0.0001
FR096	활동성	총자본회전을	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

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

<선형결합 회귀식>

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

<로짓 함수>

<다변량 로짓회귀분석 추정 결과>

III. 분석

06. 논문 결과 분석

다변량분석결과

Odds 추정 결과

변수코드	범주	변수	Odds	95% Wald 신뢰구간	
FR010	부채상환능력	금융비용대 부채비율	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	현금흐름대총자산비율	1.012	1.006	1.019
FR046	수익성	영업이익대총자산비율	0.986	0.974	0.999
FR065	안정성	유보이익대총자산비율	1.021	1.011	1.031
FR076	안정성	자기자본비율	0.972	0.961	0.983
FR080	활동성	자본금회전을	0.984	0.973	0.996
FR090	안정성	차입금의존도	0.983	0.976	0.990
FR092	활동성	총부채회전을	0.225	0.143	0.353
FR095	생산성	총자본투자효율	0.945	0.929	0.962
FR096	활동성	총자본회전을	4.359	2.615	7.266
FR097	수익성	총자산순이익률	1.041	1.020	1.062
FR106	유동성	현금성자산비율	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

〈각 재무비율 수들에 대한 odds 추정치〉

III. 분석

06. 논문 결과 분석

다변량분석결과

분류정확도, 예측력

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

<다변량 로짓모형의 분류정확도와 예측력 지표>

Concordant(%)	1인 pairs 개수 / 총 pairs 개수
Discordant(%)	0인 pairs 개수 / 총 pairs 개수
Tied(%)	0.5인 pairs 개수 / 총 pairs 개수
Pairs	우량집단 개수 x 불량집단 개수

0 => (우량 p > 불량 p) 경우

0.5 => (우량 p = 불량 p) 경우

1 => (우량 p < 불량 p) 경우

<결과 값 산출 방법>

III. 분석

06. 논문 결과 분석

다변량분석결과

부도 확률을 등급으로 환산한 결과

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%

등급구분 기준 :
금융감독원(2005)
신용리스크 내부등급 기본 세부지침

10개 등급이긴 하지만 10등급은
실제 부도발생 기업이라 제외

<예상부도확률에 따른 등급구분과 등급별 부도율>

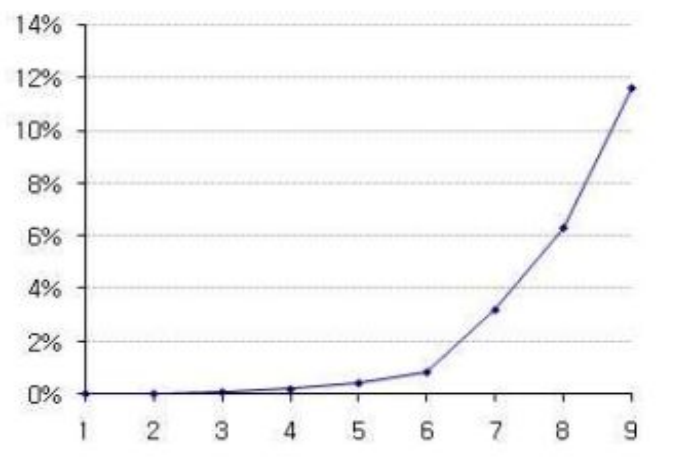
III. 분석

06. 논문 결과 분석

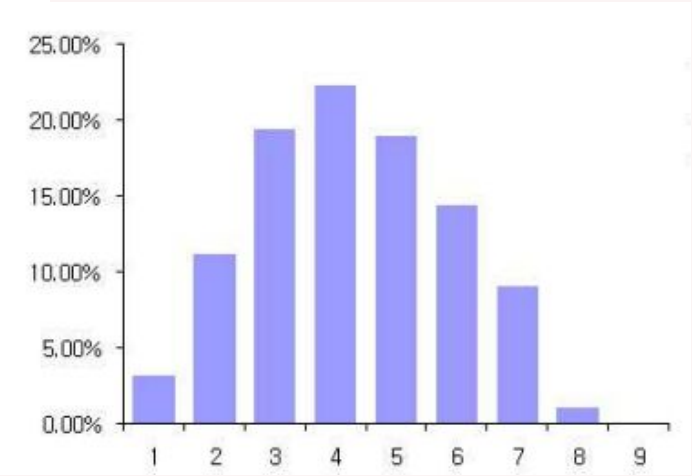
다변량분석결과

부도 확률을 등급으로 환산한 결과

등급	표본기업수			등급분포(비율, %)			누적분포(비율, %)			AU ROC
	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%			



〈등급별 부도율〉



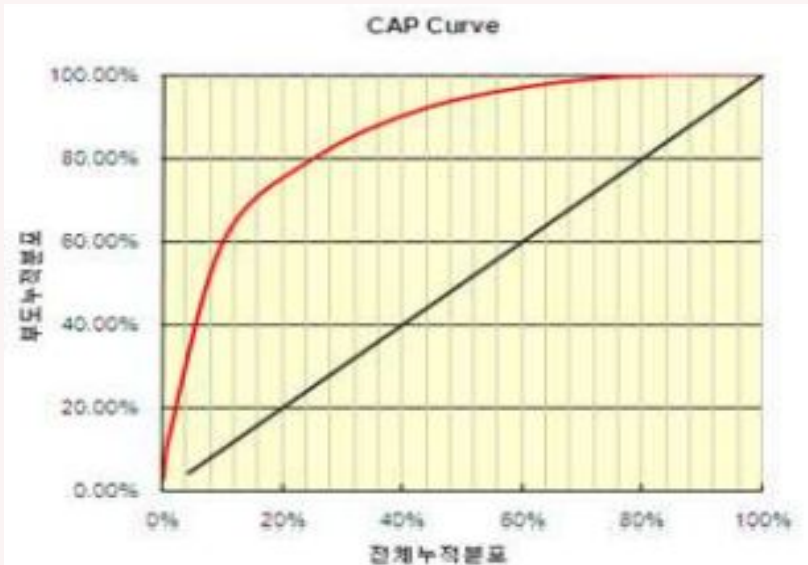
〈등급별 차주 분포〉

III. 분석

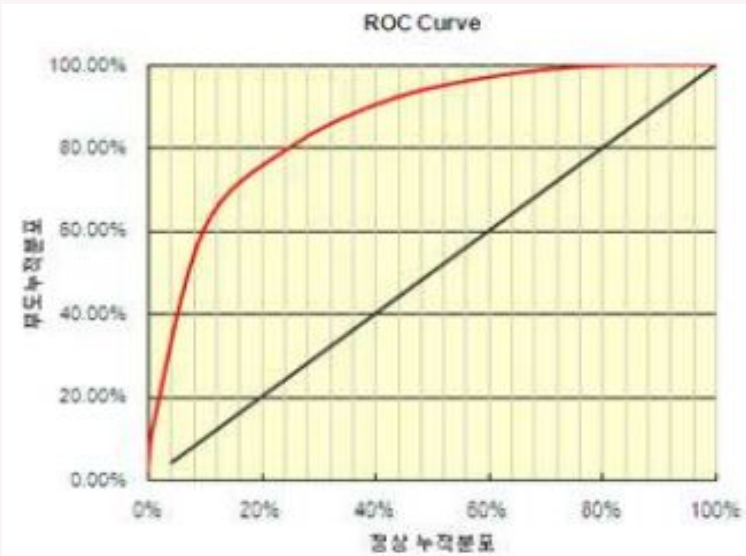
06. 논문 결과 분석

다변량분석결과

부도 확률을 등급으로 환산한 결과



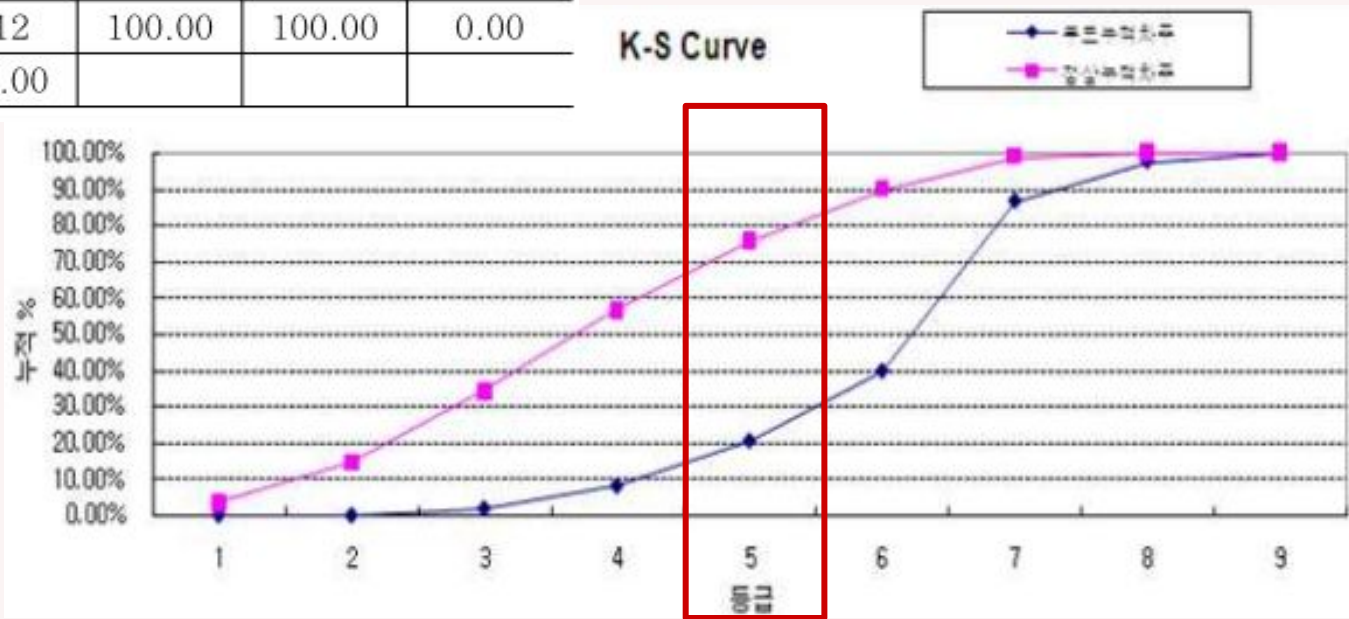
<등급화된 모형의 CAP 곡선>



<등급화된 모형의 ROC 곡선>

등급	등급분포비율(%)		누적분포비율(%)		K-S
	부도	정상	부도	정상	
1	0.00	3.22	0.00	3.22	3.22
2	0.00	11.29	0.00	14.51	14.51
3	1.86	19.60	1.86	34.10	32.24
4	6.21	22.40	8.07	56.51	48.43
5	12.42	19.03	20.50	75.53	55.04
6	19.25	14.46	39.75	89.99	50.24
7	46.89	8.88	86.65	98.87	12.23
8	10.87	1.01	97.52	99.88	2.37
9	2.48	0.12	100.00	100.00	0.00
합계	100.00	100.00			

<등급화된 모형의 K-S 통계량>



III. 분석

06. 논문 결과 분석

다변량분석결과

극단치와 결측치를 조정에 따른 결과 비교

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

<극단치와 결측치를 조정한 경우>



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

<극단치와 결측치를 조정하지 않은 경우>

극단치/결측치 조정을 안 한 결과 또한 준수한 모델 성능

III. 분석

06. 논문 결과 분석

다변량분석결과

연도별 / 업종별로 나누어 모델링 재진행한 결과

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%

<연도별 변별력 분석 결과(2003년 ~ 2006년)>

➡ 연도는 부도예측에 큰 영향이 없음을 알 수 있음.

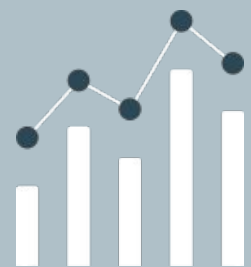
구분	제조업	건설업	도소매업	서비스업	전체
AR	72.8%	52.5%	71.6%	54.1%	70.0%
K-S	58.2%	40.9%	61.4%	39.4%	55.04%

<업종별 변별력 분석 결과>

➡ 건설업과 서비스업 부도예측이 제조업과 도소매업에서의 부도예측보다 비해 효과적이지 않았음.

IV. 고찰

01.기대효과



시계열적 강건한 결과

극단치, 결측치, 연도에 대한 영향 미비



부도예측에 대한 유용성

한국시장의 외부감사
대상기업의 부도예측에 유용



산업별 특성 반영

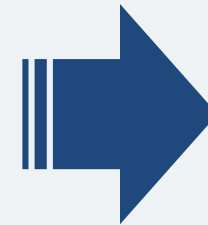
재무비율 부도예측모형의 구축에
산업별 특성 반영에 대한 필요성 제시

본 연구에서 보인 산업별 특성의 차이를 반영한 부도예측 모형의 구축과 평가는
실무적으로 의미있는 결과를 제공

IV. 고찰

02.한계점/ 향후 연구방향

외부감사대상기업 전체에
적용가능한 모형의 구축에 초점을
맞추어 재무비율만을 정보로 사용



주가 등 시장정보와 질적인 요인들을
다양하게 반영하는
연구로의 확대가 필요

금융결제원의 당좌거래정보만을
기초한 부도기업 파악

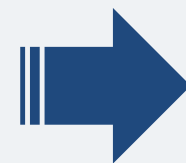


은행연합회의 자료를 이용하여 모형의
질적 개선이 필요

V. 향후 계획

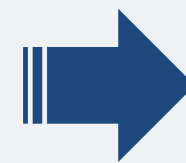
적용방안

부도의 정의 및
데이터 수집, 전처리 방법 활용



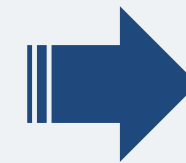
충분한 토의과정, EDA를 통해
최적의 방법 도출 후 적용

추후 부도모형 생성을 위한
Analysis Flow 활용



Flow 구조를 참고하여
다양한 통계적 기법 적용
(머신러닝 모델 추가 고려)

업종 별 부도 예측 정확도의
차이가 존재



특정 업종을 채택에 대한
연구 필요

THANK YOU