

학자금 대출 연체의 신용위험 평점 모형 개발

한준태¹ · 정진아²

¹²한국장학재단 통계분석팀

접수 2016년 7월 26일, 수정 2016년 9월 9일, 게재확정 2016년 9월 19일

요약

본 연구는 한국장학재단 일반상환 학자금 대출 연체자를 대상으로 연체 미회수 그룹으로 분류될 수 있는 위험요인들을 파악하고, 학자금 대출 연체 회수 예측모형을 개발하였다. 또한 개발된 예측모형을 활용하여 그에 따른 신용위험 평점표를 작성하였다. 예측모형 개발은 연체기간에 따라 총 3가지 모형 (Model 1: 연체 1개월 모형, Model 2: 연체 2개월 모형, Model 3: 연체 3개월 이상 모형)으로 로지스틱 회귀분석 분석을 적용하였다. 연체기간 구분은 금융권에서 일반적으로 사용하고 있는 연체회수모형의 단위를 준용하여 1개월 단위를 기준으로 연체 1개월, 연체 2개월, 연체 3개월 이상으로 구분하였다. 연체 1개월 모형 (Model 1)에서는 연체계좌수, 이체일자, 연체잔액, 소득분위가 영향력이 큰 것으로 나타났으며, 연체 2개월 모형 (Model 2)에서는 연체 일수, 연체잔액, 이체일자, 연체금액이 중요한 것으로 나타났다. 마지막으로 연체 3개월 이상 모형 (Model 3)에서는 최근 3개월 이내 연체 횟수, 이체일자, 연체계좌수, 연체액의 영향력이 큰 것으로 나타났다. 본 연구에서 개발된 연체 회수 모형이나 평점표를 바탕으로 연체 채권관리를 함에 있어 좀더 세분화된 관리서비스를 제공하고, 상담센터의 상담원이 연체자의 평점에 따라 상담전략을 세울 수 있는 기초자료가 될 수 있을 것으로 사료된다.

주요용어: 로지스틱 회귀분석, 신용위험 평점표, 학자금대출.

1. 서론

1.1. 연구 배경

고등교육 수요가 급증하면서 한국의 고등교육은 선택이 아닌 필수로 인식되고 있다. 70%를 넘어서는 높은 대학진학률은 고등교육의 국민적 열망을 직접적으로 나타낸다고 볼 수 있다. 학자금지원 사업은 이처럼 높은 국민적 교육열을 장려하기 위해 도입되었다. 2009년 한국장학재단의 설립으로 정부주도의 본격적인 학자금 대출 지원 사업이 실시되었으며, 높은 교육비를 요구하는 고등교육에 저금리를 통한 자본조달은 고등교육의 실질적 기회 확대라는 목표를 달성할 수 있게 해주었다. 하지만 정부주도의 학자금사업이 실시된 이후 늘어난 가계부채와 실질적 소득이 없는 대학생에게 발생한 부채는 간과할 수 없는 사회적 문제로 발전하게 되었다.

우리보다 학자금 지원 사업을 먼저 시작한 미국의 경우, 경제회복의 큰 장애물 중 하나로 학자금 대출금이 지목되고 있으며, 청년시절 사용한 학자금 대출금을 상환하지 못해 65세 이상까지 경제적 어려움을 겪는 노령인구가 늘어나고 있다. 이와 같이 학자금 대출사업은 개인의 경제활동에 직접적 영향을 미

¹ 교신저자:(41200) 대구광역시 동구 신암로 125, 한국장학재단 통계분석팀, 팀장.

E-mail: hanjt@kosaf.go.kr

² (41200) 대구광역시 동구 신암로 125, 한국장학재단 통계분석팀, 주임.

칠 뿐 아니라 국가경제와도 밀접하게 연결되어 있으며, 앞으로 국내에서도 학자금 대손, 미상환 문제가 나타날 수 있다.

한국장학재단은 고등교육에 대한 접근성 확보를 지원하기 위하여 2009년 설립 이후 대학생과 대학원생을 대상으로 등록금 (tuition) 및 생활비 (living expenses)를 장기 저리로 융자하는 학자금 대출제도를 운영하고 있다. 등록금에는 입학금, 수업료 등이 포함되며 생활비에는 숙식비, 교재구입비, 교통비 등 생활안정을 위한 일체의 비용이 포함되어 있다. 현재 한국장학재단이 제공하고 있는 학자금 대출 제도는 취업 후 상환 학자금 대출, 일반상환 학자금 대출, 농어촌 출신 대학생 학자금 대출까지 총 세 종류가 있으며, 대출제도 간에는 대출 대상 및 대출 조건, 상환 방식 등에서 차이가 존재한다.

취업 후 상환 학자금 대출은 2년제 및 4년제 대학 재학 중인 학부생을 대상으로 재학 중 등록금과 생활비를 대출하고 졸업 후 일정 수준 이상의 소득이 발생하는 경우 원리금을 상환하는 것으로 졸업 후 일정 수준 이상의 소득이 생기는 경우에만 상환의무가 발생하는 특징에 착안하여 ‘소득연계대출 (income contingent loan; ICL)’로 부르기도 한다.

일반상환 학자금 대출은 소득 8분위 이상의 학부생과 대학원생을 대상으로 등록금과 생활비를 대출하고 미리 정해진 방식에 따라 원리금이 상환되는 학자금 대출 제도로 채무자가 거치기간과 상환기간, 상환방식을 미리 선택하고 그에 따라 상환의무가 발생하는 일반적인 형태의 대출 제도이다. 취업 후 상환 학자금 대출 대상에서 제외되는 소득 8분위 이상의 고소득 가구에 속하는 학부생과 대학원생의 학자금을 장기 저리로 대출이 가능하다.

한국장학재단에는 학자금 대출로 금융채무 불이행자가 된 학생들을 위한 구제제도가 있다. 한국장학재단의 학자금 대출은 시중은행과는 달리 ‘6개월 이상 장기 연체자’에 대해서만 금융채무 불이행자로 등록하고 있으며, 2016년 7월부터는 금융채무 불이행자가 연체된 대출 잔액을 상환하는 경우 금융채무 불이행 기록보존 기간을 5년에서 1년으로 단축하는 등 학자금 대출을 받은 학생들의 상환 부담을 최대한 낮췄다. 이외에도 금융채무 불이행자 회복지원 제도로 분할상환 제도, 손해금 감면 제도, 금융채무 불이행자 신용회복 제도, 취업 연계 신용회복 지원 제도 등을 운용 중이다.

연체 금액만을 연도별로 보면, 일반학자금 대출이 처음 시행된 2009년에는 400억으로 시작하여 다음 해 약 1.2조, 2011년에는 큰폭으로 올라 약 2조 9천억 원이 연체되었다. 2012년 약 3조 7천억 원을 기점으로 연체금액은 조금씩 감소하고 있는 추세이다. 연체율 역시 3.4%를 시작으로 2012년 8.5%의 최고 연체율을 갱신한 후부터 7~8%대를 유지하고 있다. 연체 현황을 보았을 때 한번 높아진 연체율은 쉽게 떨어지지 않기 때문에 연체채권에 대한 별도 관리가 반드시 필요한 것으로 보인다.

1.2. 선행 연구

Khieu 등 (2012)은 Moody's Ultimate Recovery Database를 이용하여, 은행 대출 회수율에 대출 특성 (대출규모, 대출 종류, 담보), 대출자의 특성, 기업 특성, 경제상태 등의 변수가 미치는 영향을 연구하였는데, 대출의 특성 (대출규모, 대출 종류, 담보)이 대출자의 특성보다 회수율에 좀 더 영향을 미치는 요인으로 조사되었다. Grunert와 Weber (2009)은 회수율의 분포와 담보물, 대출자의 신용도, 회사 규모, 고객과의 관계 등을 분석하였으며, 담보가 많을수록 회수율이 높아지지만, 대출자 및 회사 규모가 위험요소가 있을 경우 회수율에 나쁜 영향을 미치는 것을 확인하였다. Zurada와 Zurada (2002)는 고객이 채무 불이행할지 또는 대출 상환할지를 예측하기 위해 세가지 의사결정나무 분석 알고리즘 (카이제곱, 엔트로피, 지니계수)의 효율성을 분석하고 비교하였다. 이밖에 기업의 부도 예측에 관한 다양한 방법들은 연구되어 왔다 (Altman 등, 1977; Pompe와 Bilderbeek, 2005).

최근에 Leow와 Crook (2016)은 영국 주요 은행의 2002~2011년까지의 신용카드 대출 자료를 이용하여 신용카드 계좌의 부도를 예측하기 위한 모형 개발을 연구하였는데, 금융위기가 있었던 2008년을

기점으로 2002~2007년과 2008~2011년의 두 개의 데이터 셋을 나누어 생존분석 모형을 적용한 결과 각각의 회귀식 추정계수 및 주요요인 결과가 다소 상이하게 나타남을 확인하였다.

국내에서의 연구로는 Jin 등 (2012)의 연구에서 신용평점표 작성시 각 특성변수들을 몇 개의 속성들로 나눌 때 기준이 되는 절단값을 별점화 기반으로한 절단값 선택법을 제안하였다. Kim과 Kim (2013)은 국내 여신기관의 자료를 사용하여 대출심사에 관한 연구를 진행하였으며, 모형구축에 있어서 안정적이고 정확한 예측을 보이는 모형을 찾기 위해 데이터마이닝 기법 등을 사용하여 다양한 모형을 구축하고 비교하였으며, Woo 등 (2013)은 신용 리스크를 최소화하는 도구 중 하나인 신용 평점 모형을 목표변수가 순서형 다항자료 또는 경시적 이항 자료 같은 형태를 고려한 신용 평점 모형구축 방법을 제시하였다.

위에서 언급한 문헌들은 은행이나 카드사 등 일반적인 대출에서의 연구들이 대부분이다. 학자금 지원 사업의 경우 매년 막대한 정부예산이 투입되고 한국장학재단이 관리·운영하는 규모가 매년 증가함에도 불구하고 학자금 대출에서의 연체 채권회수 관련 국내외 문헌을 좀처럼 발견하기 어렵다.

따라서 한국장학재단 학자금 대출의 채권회수에서 주요 요인 간에 어떤 연관성이 있는지에 대한 실증적 연구의 필요성이 대두되었다. 이러한 요구에 발맞춰 대출 연체자의 다양한 정보를 기반으로 한 고객 세분화 (Customer segmentation)를 통하여 연체를 효율적으로 방지하고 조기에 채권을 정리할 수 있는 적절한 채권회수 모형이 개발되어야 할 것이다. 향후, 이러한 예측 및 관리모형은 학자금 대출에서의 연체자에 대한 효율적인 관리와 적절한 채권관리 방안 등에 활용될 수 있을 것이며 정부 학자금대출 사업의 선순환구조를 확립하는데 기여할 것이다.

재단의 학자금대출은 요건대출로 사전 심사 없이 대출이 가능하기 때문에 대출 이후 관리가 더 중요하다. 또한 기존의 연구된 채권회수모형의 경우에는 금융기관 (카드사, 은행)에 최적화 되어있기 때문에 재단 대출사업에 그대로 적용하기에는 어려움이 있다. 따라서 ‘학생을 대상으로 하는 요건대출’이라는 학자금 대출의 특성을 반영한 채권회수모형이 필요하다.

앞서 연구한 ‘부실 고위험군 예측모형 개발’ (Choi 등, 2015)에서는 채권관리 단계에서 장기연체가 될 수 있는 고위험군 모형을 선별적으로 관리 할 수 있도록 기초자료로써 활용하였다. 그러나 정책금융적 성격의 재단 채권은 적극적인 채권관리의 어려움이 존재하기 때문에 채권회수 모형 등을 통한 연체 채권의 효율적인 회수 방안으로 활용할 필요가 있다.

본 연구에서는 재단 학자금 대출의 특성과 연체자의 개인·금융 정보를 반영한 연체 회수모형을 구축하고, 개발된 예측모형을 활용하여 그에 따른 신용위험 평점표를 작성하여 연체기간에 따라 연체 채권관리 방법을 세분화하여 기간별로 관리할 수 있는 모형을 제안하고자 한다.

본 연구의 2장에서는 모형 개발을 위한 자료의 설명과 분석 방법에 대해 소개하였다. 3장에서는 개발된 예측모형의 결과와 개발된 모형의 평가와 평점표 작성에 대해 다루었으며, 4장에서는 결론 및 향후 연구방향을 제시하였다.

2. 모형 개발

본 연구에서 개발하고자 하는 학자금 대출 연체 회수 예측모형 (the debt collection predictive model)은 한국장학재단의 일반상환 학자금대출 연체자 중에서 회수 고객과 미회수 고객을 미리 예측하는 모형이며, 개발된 예측모형을 활용하여 그에 따른 신용위험 평점표를 작성하였다.

연체 회수 모형 개발을 위한 자료는 한국장학재단의 원천시스템 및 데이터웨어하우스 (data warehouse)에서 최근 3개월 기준 (2015.9월~11월)으로 일반상환 학자금대출 연체자의 인구사회학적 특성, 소득분위, 학교 및 학교개황, 대출정보 자료, 국가장학금 수혜 정보 등을 이용하였다. 분석 패키지는 SAS 9.4를 이용하였다.

2.1. 자료 설명

연체채권에 영향을 주는 요인은 관련 연구자료 및 관련 전문가 자문 등을 통하여, 인구사회학적 특성(연령, 성별, 결혼여부), 소득분위, 국가장학금 수혜 유무, 대출실행 당시의 학교개황 정보로 대학구분(학부, 대학원), 대학설립 형태(국공립, 사립), 대학 소재지(수도권, 비수도권), 학교개황 정보 중 계열정보(인문계열, 사회계열, 교육계열, 공학계열, 자연계열, 의약계열, 예체능계열), 대출정보(대출 거치기간, 월평균 상환액, 대출금리, 이체일자), 연체관련 정보(연체계좌 수, 연체미납금, 연체잔액, 과거 연체경험 유무, 연체일수)를 설명변수로 포함하였다. 여기서 월평균 상환액은 대출계좌별로 거치기간과 대출기간을 최댓값으로 설정하여 총 대출금액을 대출기간에서 거치기간을 제외한 개월 수로 나눈 것을 월평균 상환액으로 산출하였다.

연체 회수 모형의 종속변수(목표변수)는 연구대상자로 M월(9월, 10월, 11월)에 발췌된 연체자가 익월(M+1월: 10월, 11월, 12월)에 연체금액을 완전해소 한 경우를 회수(good)로 정의하였으며, 그렇지 못한 경우를 미회수(bad)로 정의하였다. Table 2.1은 본 연구에 활용된 변수들의 설명이다.

Table 2.1 Data description in the analysis

Variable	Definition
Bad (target variable)	Person who doesn't repay the debt (1: Yes, 2: NO)
Sex	Male, Female
Age	≤20, 21-23, 24-26, 27-30, ≥31
Marital status	Married, Not married
Household income	(Model 1) 1:Class 0, 2:Class 1, 3:Class 2-4, 4:Class 5-6, 5:Class 7-9, 6:Class10, 7:etc (Model 2) 1:Class 0, 2:Class 1, 3:Class 2-4, 4:Class 5-7, 5:Class 8-9, 6:Class10, 7:etc (Model 3) 1:Class 0, 2:Class 1, 3:Class 2-4, 4:Class 5-7, 5:Class 8-9, 6:Class10, 7:etc
National grant	No, Yes
School foundation	National, Private
Location	Metropolitan, Local
Level of schooling	Graduate, Undergraduate
Field of study	Humanities, Social Sciences, Education, Engineering, Natural Sciences, Medical Science & Pharmacy, Art & Physical Education
Grace period (Unit: months)	(Model 1) 1: ≤119 2: ≥120 (Model 2) 1: ≤12, 2: 13-36, 3: 37-48, 4: 49-60, 5: 61-72, 6: ≥73 (Model 3) 1: ≤24, 2: 25-48, 3: 49-72, 4: 73-96, 5: 97-120, 6: 121-168 7: ≥169
Monthly repayment	Amount of loans / (Loan period - Grace period)
Interest rate(%)	2.7, 2.9, 3.9, 4.9, 5.2, 5.7, 5.8
Due date for payment	≤5, 6-10, 11-15, 16-20, 21-25, ≥26
Overdue account	(Model 1) 1: 1, 2: 2, 3: 3-4, 4: 5, 5: 6-9, 6: ≥10 (Model 2) 1: 1, 2: 2, 3: 3-7, 4: 8-9, 5: ≥10 (Model 3) 1: 1, 2: 2, 3: 3-7, 4: 8-9, 5: ≥10
Arrears (Unit: KRW thousands)	(Model 1) 1: <100, 2: 100-200, 3: 200-300, 4: ≥300 (Model 2) 1: <15, 2: 15-30, 3: 30-60, 4: 60-130, 5: 130-260, 6: 260-680, 7: ≥680 (Model 3) 1: <45, 2: 45-100, 3: 100-270, 4: 270-500, 5: 500-960, 6: 960-1700, 7: ≥1700
Balance (Unit: KRW thousands)	(Model 1) 1: <1200, 2: 1200-2200, 3: 2200-3400, 4: 3400-6000, 5: 6000-13000, 6: ≥13000 (Model 2) 1: <1000, 2: 1000-1600, 3: 1600-2800, 4: 2800-4000, 5: 4000-7500, 6: ≥7500 (Model 3) 1: <1000, 2: 1000-1600, 3: 1600-2800, 4: 2800-3300, 5: 3300-12000, 6: ≥12000
Overdue record	Yes, No
Days in arrears	(Model 1) 1: ≤10, 2: 11-28, 3: 29, 4: ≥30 (Model 2) 1: ≤35, 2: 36-40, 3: 41-45, 4: 46-50, 5: ≥51 (Model 3) 1: ≤72, 2: 73-94, 3: 95-104, 4: 105-121, 5: 122-134, 6: 135-154, 7: ≥155

2.2. 개발 프로세스

한국장학재단 일반상환 학자금 대출자 중 최근 3개월 기준(2015.9월-11월) 연체자를 대상으로 모형을 구축하였으며, 연체기간이 지날수록 채권의 속성이 변한다라는 가정하에 3개월간(9월, 10월, 11월) 각각 연체 기간에 따라 연체 1개월, 연체 2개월, 연체 3개월 이상으로 대상자를 발췌한 다음 총 3가지 모형(Model 1: 연체 1개월 모형(N=3,457명), Model 2: 연체 2개월 모형(N=53,654명), Model 3: 연체 3개월 이상 모형(N=64,092명))을 개발하였다. 분석패키지는 SAS 9.4를 이용하였으며, 로지스틱 회귀 분석을 적용하여 진행하였다.

본 연구에서 활용한 변수는 Table 2.1의 변수를 사용하였으며, Model별로 연체잔액 대비 연체액, 최근 3개월내 연체횟수 등의 파생변수를 활용하였다. 그리고 모형별로 독립변수의 범주화는 Information value (IV)를 활용하여 IV가 최대가 되는 범주를 각각 적용하였다. IV는 독립변수와 이분형 종속변수 (불량집단 및 우량집단) 간의 구성비를 이용한 것으로 일정한 공식에 대입하여 구할 수 있으며 식 (2.1)과 같다 (Siddiqi, 2006).

$$IV = \sum_{k=1}^K \left[(\%good_k - \%bad_k) \times \log \left(\frac{\%good_k}{\%bad_k} \right) \right] \quad (2.1)$$

여기서, 독립변수 범주의 수인 K 는 $K \geq 2$ 이다.

본 연구에서는 모형별로 각각 범주 및 재그룹하여 독립변수를 설정하여 모형개발에 적용하였다.

분석데이터는 크게 분석용 (training data)과 평가용 (validation data)으로 구분하였고, 분석용과 평가용 데이터는 일반상환학자금 대출 연체자를 기준으로 6 대 4의 비율로 분할하여 생성하였다. 또한, 개발된 예측모형의 성능에 대한 평가는 ROC (receiver operation characteristic) 곡선 및 오분류율을 이용하였다.

2.3. 주요 요인의 일반적 현황

연구대상자 총 121,203명 중 회수 (good)는 18,076명, 미회수 (bad)는 103,127명으로 회수 (good)는 약 14.9%, 미회수 (bad)는 약 85.1%로 나타났으며, Model 1에서의 회수 (good)가 28.8%인 반면 Model 3에서는 9.5%로 연체기간이 길수록 회수 (good) 그룹의 비율이 낮아짐을 알 수 있다 (Table 2.2).

Table 2.2 Target variable for the data set

	Good	Bad	Total
	N (%)	N (%)	N (%)
Model 1	989 (28.8%)	2,468 (71.4%)	3,457 (100.0%)
Model 2	11,008 (20.5%)	42,646 (79.5%)	53,654 (100.0%)
Model 3	6,079 (9.5%)	58,013 (90.5%)	64,092 (100.0%)
Total	18,076 (14.9%)	103,127 (85.1%)	121,203 (100.0%)

각 설명변수의 현황분석을 실시하기 위해 회수 (good)와 미회수 (bad) 그룹간의 교차분석을 Model별로 수행하였다 (Table 2.3).

Model 1에서는 연령, 소득분위, 대학구분 (학부, 대학원), 거치기간, 월평균 상환액, 대출금리, 이체일자, 연체계좌 수, 연체잔액, 연체일수에서 통계적으로 유의한 차이를 보였으나, Model 2에서는 성별, 연령, 결혼유무, 소득분위, 국가장학금 수혜유무, 대학구분 (학부, 대학원), 대학 소재지, 학과 계열구분, 거치기간, 월평균 상환액, 대출금리, 이체일자, 연체계좌수, 연체액, 연체잔액, 연체일수에서 통계적으로 유의한 차이를 보였다. Model 3에서는 거의 모든 변수에서 통계적인 차이를 보였지만, 결혼유무, 국가장학금 수혜유무, 대학구분 (학부, 대학원), 대학 소재지에서는 회수와 미회수 그룹간의 통계적인 차이는 없었다.

주요 요인의 일반적 현황을 확인한 결과 Model별로 주요 요인이 조금씩 다른것을 확인 할 수 있었으며, 특히 Model 1에서는 학교 및 학과 관련된 정보 즉, 대학설립 형태 (국공립, 사립), 대학 소재지 (수도권, 비수도권), 대학구분 (학부, 대학원), 학과 계열구분 (인문계열, 사회계열, 교육계열, 공학계열, 자연계열, 의약계열, 예체능계열)의 정보는 주요 요인으로 나타나지 않았다. 반면, 연령, 소득분위, 이체일자, 연체계좌수, 연체잔액, 연체일수 등은 모든 Model에서 회수 (good)와 미회수 (bad) 그룹간에 통계적으로 차이가 나타났다.

Table 2.3 Descriptive characteristics for the variables in the analysis

		Model 1			Model 2			Model 3		
		Good N(%)	Bad N(%)	χ^2	Good N(%)	Bad N(%)	χ^2	Good N(%)	Bad N(%)	χ^2
Sex	Male	485 (49.0)	1,210 (49.0)	0	3,098 (51.0)	28,197 (48.6)	12.24 **	5,102 (46.3)	21,176 (49.7)	38.29 **
	Female	504 (51.0)	1,258 (51.0)		2,981 (49.0)	29,816 (51.4)		5,906 (53.7)	21,470 (50.3)	
Age	≤ 20	188 (19.0)	597 (24.2)	19.68 **	1,761 (29.0)	17,267 (29.8)	36.72 **	2,624 (23.8)	11,751 (27.6)	67.20 **
	21-23	236 (23.9)	631 (25.6)		1,638 (26.9)	14,657 (25.3)		2,994 (27.2)	11,155 (26.2)	
	24-26	243 (24.6)	582 (23.6)		1,329 (21.9)	11,800 (20.3)		2,599 (23.6)	9,303 (21.8)	
	27-30	197 (19.9)	376 (15.2)		730 (12.0)	7,045 (12.1)		1,551 (14.1)	5,610 (13.2)	
	≥ 31	125 (12.6)	282 (11.4)		621 (10.2)	7,244 (12.5)		1,240 (11.3)	4,827 (11.3)	
Marital status	Not married	909 (91.3)	2,267 (91.9)	0.28	5,663 (93.2)	52,913 (91.2)	26.54 **	10,133 (92.1)	39,216 (92.0)	0.10
	Married	86 (8.7)	201 (8.1)		416 (6.8)	5,100 (8.8)		875 (7.9)	3,430 (8.0)	
Household income†	1	21 (2.1)	93 (3.8)	30.33 **	281 (4.6)	2,960 (5.1)	27.44 **	319 (2.9)	1,756 (4.1)	102.61 **
	2	37 (3.7)	127 (5.1)		313 (5.1)	2,834 (4.9)		457 (4.2)	2,135 (5.0)	
	3	153 (15.5)	457 (18.5)		1,263 (20.8)	13,279 (22.9)		1,902 (17.3)	8,330 (19.5)	
	4	61 (6.2)	133 (5.4)		678 (11.2)	6,844 (11.8)		1,154 (10.5)	4,382 (10.3)	
	5	225 (22.8)	653 (26.5)		1,472 (24.2)	12,896 (22.2)		2,844 (25.8)	10,119 (23.7)	
	6	110 (11.1)	227 (9.2)		515 (8.5)	4,715 (8.1)		1,114 (10.1)	3,800 (8.9)	
	7	382 (38.6)	778 (31.5)		1,557 (25.6)	14,485 (25.0)		3,218 (29.2)	12,124 (28.4)	
National grant	No	706 (71.4)	1,751 (70.9)	0.06	4,199 (69.1)	42,109 (72.6)	33.84 **	7,352 (66.8)	28,548 (66.9)	0.09
	Yes	283 (28.6)	717 (29.1)		1,880 (30.9)	15,904 (27.4)		3,656 (33.2)	14,098 (33.1)	
Level of schooling	Graduate	164 (16.6)	336 (13.6)	5.02 *	1,871 (17.0)	5,884 (13.8)	72.43 **	812 (13.4)	7,385 (12.7)	1.94
	Undergraduate	825 (83.4)	2,132 (86.4)		9,137 (83.0)	36,762 (86.2)		5,267 (86.6)	50,628 (87.3)	
Location	Metropolitan	177 (17.9)	465 (18.8)	0.41	2,192 (19.9)	7,722 (18.1)	18.93 **	1,052 (17.3)	10,547 (18.2)	2.84
	Local	812 (82.1)	2,003 (81.2)		8,816 (80.1)	34,924 (81.9)		5,027 (82.7)	47,466 (81.8)	
School foundation	National	75 (7.6)	198 (8.0)	0.18	1,305 (11.9)	5,153 (12.1)	0.43	707 (11.6)	6,057 (10.4)	8.24 **
	Private	914 (92.4)	2,270 (92.0)		9,703 (88.1)	37,493 (87.9)		5,372 (88.4)	51,956 (89.6)	
Field of study	Humanities	146 (14.8)	374 (15.2)	4.62	1,740 (15.8)	6,919 (16.2)	30.53 **	873 (14.4)	8,925 (15.4)	19.79 **
	Social Sciences	221 (22.3)	583 (23.6)		2,448 (22.2)	9,889 (23.2)		1,375 (22.6)	13,748 (23.7)	
	Education	49 (5.0)	103 (4.2)		496 (4.5)	1,932 (4.5)		252 (4.1)	2,418 (4.2)	
	Engineering	223 (22.5)	505 (20.5)		2,214 (20.1)	9,053 (21.2)		1,396 (23.0)	12,091 (20.8)	
	Natural Sciences	111 (11.2)	317 (12.8)		1,391 (12.6)	5,195 (12.2)		741 (12.2)	7,065 (12.2)	
	Medical	36 (3.6)	84 (3.4)		436 (4.0)	1,693 (4.0)		241 (4.0)	2,130 (3.7)	
	Science	203 (20.5)	502 (20.3)		2,283 (20.7)	7,965 (18.7)		1,201 (19.8)	11,636 (20.1)	
	Arts & Physical	98 (9.9)	262 (10.6)		1,152 (10.5)	4,302 (10.1)		676 (11.1)	5,328 (9.2)	
	Other	104 (10.5)	218 (8.8)		1,034 (9.4)	3,528 (8.3)		544 (8.9)	4,642 (8.0)	
	Other	155 (15.7)	344 (13.9)		1,528 (13.9)	5,959 (14.0)		890 (14.6)	8,728 (15.0)	
Monthly repayment	< 28000	60 (6.1)	285 (11.5)	29.02 **	951 (8.6)	5,084 (11.9)	120.62 **	680 (11.2)	4,918 (8.5)	252.46 **
	28000-50000	179 (18.1)	477 (19.3)		2,083 (18.9)	8,412 (19.7)		1,221 (20.1)	9,552 (16.5)	
	50000-65000	98 (9.9)	262 (10.6)		1,152 (10.5)	4,302 (10.1)		676 (11.1)	5,328 (9.2)	
	65000-80000	104 (10.5)	218 (8.8)		1,034 (9.4)	3,528 (8.3)		544 (8.9)	4,642 (8.0)	
	80000-110000	155 (15.7)	344 (13.9)		1,528 (13.9)	5,959 (14.0)		890 (14.6)	8,728 (15.0)	
	110000-250000	282 (28.5)	643 (26.1)		3,098 (28.1)	10,983 (25.8)		1,574 (25.9)	17,264 (29.8)	
	≥ 250000	111 (11.2)	239 (9.7)		1,162 (10.6)	4,378 (10.3)		494 (8.1)	7,581 (13.1)	

*p < 0.05, **p < 0.01, †: see Table 2.1

Table 2.3 Continued

		Model 1			Model 2			Model 3		
		Good N(%)	Bad N(%)	χ^2	Good N(%)	Bad N(%)	χ^2	Good N(%)	Bad N(%)	χ^2
Interest rate	2.7	54 (5.5)	154 (6.2)	33.33 **	309 (2.8)	881 (2.1)	65.06 **	30 (0.5)	158 (0.3)	89.89 **
	2.9-3.9	504 (51.0)	1,045 (42.3)		5,255 (47.7)	19,562 (45.9)		2,601 (42.8)	21,973 (37.9)	
	4.9-5.2	362 (36.6)	1,057 (42.8)	3,897 (35.4)	15,202 (35.7)	2,289 (37.7)	22,360 (38.5)			
	5.7	45 (4.6)	85 (3.4)	964 (8.8)	4,115 (9.7)	654 (10.8)	7,335 (12.6)			
	5.8	24 (2.4)	127 (5.1)	583 (5.3)	2,886 (6.8)	505 (8.3)	6,187 (10.7)			
Due date for payment	5	130 (13.1)	794 (32.2)	141.74 **	894 (8.1)	6,101 (14.3)	1138.83 **	457 (7.5)	8,161 (14.1)	710.21 **
	6-10	63 (6.4)	162 (6.6)		1,442 (13.1)	8,344 (19.6)		824 (13.6)	12,137 (20.9)	
	11-15	49 (5.0)	109 (4.4)	1,295 (11.8)	6,191 (14.5)	777 (12.8)	8,150 (14.0)			
	16-20	60 (6.1)	153 (6.2)	2,298 (20.9)	8,784 (20.6)	1,415 (23.3)	12,749 (22.0)			
	21-25	92 (9.3)	149 (6.0)	2,726 (24.8)	7,766 (18.2)	1,451 (23.9)	10,132 (17.5)			
	≥ 26	595 (60.2)	1,101 (44.6)	2,353 (21.4)	5,460 (12.8)	1,155 (19.0)	6,684 (11.5)			
	Overdue account†	1	518 (52.4)	1,529 (62.0)	31.82 **	5,226 (47.5)	25,493 (59.8)	606.27 **	3,416 (56.2)	33,995 (58.6)
2		221 (22.3)	475 (19.2)	2,502 (22.7)		8,395 (19.7)	1,324 (21.8)		11,174 (19.3)	
3		143 (14.5)	284 (11.5)	2,646 (24.0)	7,014 (16.4)	1,082 (17.8)	10,318 (17.8)			
4		34 (3.4)	46 (1.9)	443 (4.0)	1,215 (2.8)	174 (2.9)	1,772 (3.1)			
5		56 (5.7)	108 (4.4)	191 (1.7)	529 (1.2)	83 (1.4)	754 (1.3)			
6		17 (1.7)	26 (1.1)							
Arrears†	1	481 (48.6)	1,257 (50.9)	4.91	672 (6.1)	4,411 (10.5)	243.65 **	928 (15.3)	5,837 (10.1)	618.40 **
	2	174 (17.6)	468 (19.0)		1,262 (11.5)	4,639 (10.9)		713 (11.7)	4,656 (8.0)	
	3	108 (10.9)	226 (9.2)	1,178 (10.7)	4,132 (9.7)	1,599 (26.3)	12,175 (21.0)			
	4	226 (22.9)	517 (20.9)	2,093 (19.0)	8,249 (19.3)	1,233 (20.3)	11,840 (20.4)			
	5			2,444 (22.2)	8,588 (20.1)	920 (15.1)	11,398 (19.6)			
	6			2,402 (21.8)	8,263 (19.4)	382 (6.3)	6,193 (10.7)			
	7			957 (8.7)	4,363 (10.2)	304 (5.0)	5,914 (10.2)			
Delinquent balance†	1	60 (6.1)	283 (11.5)	101.63 **	659 (6.0)	4,982 (11.7)	1018.81 **	536 (8.8)	7,934 (13.7)	125.76 **
	2	51 (5.2)	281 (11.4)		721 (6.5)	4,644 (10.9)		668 (11.0)	6,623 (11.4)	
	3	169 (17.1)	554 (22.4)	1,740 (15.8)	8,969 (21.0)	1,388 (22.8)	12,206 (21.0)			
	4	323 (32.7)	694 (28.1)	2,350 (21.3)	8,620 (20.2)	622 (10.2)	5,105 (8.8)			
	5	266 (26.9)	430 (17.4)	2,634 (23.9)	7,726 (18.1)	2,297 (37.8)	20,827 (35.9)			
	6	120 (12.1)	226 (9.2)	2,904 (26.4)	7,705 (18.1)	568 (9.3)	5,318 (9.2)			
Overdue record	No	832 (84.1)	2,026 (82.1)	2.04	8,733 (79.3)	34,143 (80.1)	2.89	4,583 (75.4)	35,166 (60.6)	509.81 **
	Yes	157 (15.9)	442 (17.9)		2,275 (20.7)	8,503 (19.9)		1,496 (24.6)	22,847 (39.4)	
Days in arrears†	1	60 (6.1)	94 (3.8)	178.48 **	4,663 (42.4)	10,469 (24.5)	2206.80 **	1,959 (32.2)	10,354 (17.8)	986.45 **
	2	24 (2.4)	82 (3.3)		3,067 (27.9)	10,084 (23.6)		1,129 (18.6)	11,735 (20.2)	
	3	312 (31.5)	1,375 (55.7)	1,746 (15.9)	8,050 (18.9)	794 (13.1)	5,839 (10.1)			
	4	593 (60.0)	917 (37.2)	726 (6.6)	5,377 (12.6)	464 (7.6)	6,124 (10.6)			
	5			806 (7.3)	8,666 (20.3)	555 (9.1)	5,514 (9.5)			
	6					315 (5.2)	6,194 (10.7)			
	7					863 (14.2)	12,253 (21.1)			

*p < 0.05, **p < 0.01, †: see Table 2.1

3. 연구 결과

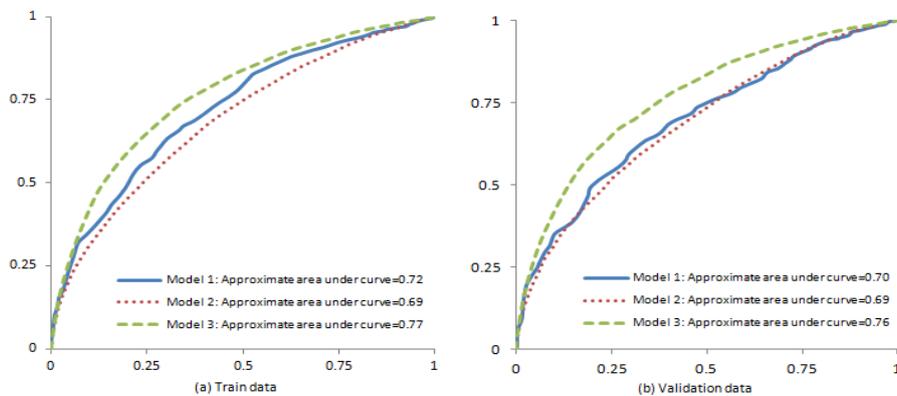
3.1. 모형의 평가

본 연구에서 개발된 모형에 대한 평가는 예측 정확도를 나타내는 평균제곱오차 (average squared error; ASE), 오분류율, Roc 곡선과 향상도에 근거하였다. 오분류율 약 0.1~0.27, ASE는 약 0.07~0.18 정도로 나타나 안정성과 정확성이 확보된 모형이 도출되었다 (Table 3.1).

Table 3.1 ASE and misclassification rate for the predictive model

	Train data: 60%		Validation data: 40%	
	Misclassification rate	ASE	Misclassification rate	ASE
Model 1:	0.27962	0.17895	0.27062	0.18559
Model 2:	0.20484	0.15129	0.20595	0.15133
Model 3:	0.09509	0.07834	0.09494	0.07840

Figure 3.1에서와 같이, Roc 곡선의 아래 면적을 나타내는 AUROC (Area under receiver operating characteristic)가 평가용 데이터에서 Model 1=0.72, Model 2=0.69, Model 3=0.77의 값을 보였으며, 검증용 데이터에서는 Model 1=0.70, Model 2=0.69, Model 3=0.76의 값을 보여 모형의 설명력과 타당성을 확보하였다.

**Figure 3.1** Roc curve for the predictive model

본 연구에서 정의된 종속변수와 독립변수 등을 고려하여 Model별로 각각 로지스틱 회귀분석을 통해 각 변수의 회귀계수 및 OR (Odds Rate)를 산출하였다 (Table 3.2).

Model 1에서는 소득분위 0분위 보다 10분위의 미회수 위험이 낮았으며, 연체계좌 수가 10개 이상일때 보다 1개일 경우가 미회수 위험이 약 3.48배 높았으며, 연체잔액이 13백만원 이상보다 6백만원~13백만원의 미회수 위험이 낮았다. 연체잔액 대비 연체금액이 3.3~15.8일 경우보다 0.4미만이거나 15.8이상일때 미회수 위험이 약 1.52배 높았으며, 연체일수에서는 10일 이하보다 29일 일때 미회수 위험이 약 2.25배, 이체일자에서는 26일 이후일때 보다 5일 이전은 약 3.35배, 6~10일은 2.51배, 11~15일은 1.83배, 16~20일은 2.79배 미회수 위험이 높게 나타났다.

Model 2를 살펴보면 남자가 여자보다 약 1.14배 미회수 위험이 높았으며, 소득분위에서는 5~10분위가 0분위보다 위험도가 낮았다. 연체계좌 수에서는 10개 이상인 경우보다 1개일 때가 약 1.88배, 2개가 약 1.41배 미회수 위험도가 높았으며, 월평균 상환액이 28,000원 미만인 경우보다 50,000~65,000원인 경우가 약 1.16배, 80,000~110,000원인 경우가 약 1.19배 위험도가 높았다. 연체액이 15,000원 미만인 경우보다 연체액이 680,000원까지는 미회수 위험도가 낮았으나, 연체잔액이 7,500,000원 이상인 경우보다 1,000,000원 미만인 경우는 약 2.50배, 1,000,000~1,600,000원은 약 2.06배, 1,600,000~2,800,000원은 약 1.76배, 2,800,000~4,000,000원은 약 1.26배 미회수 위험도가 높았고, 과거 연체 경험이 있는 경우 약 1.15배 위험도가 높았다. 연체일수 역시 많을수록 미회수 위험도가 높았으며, 국가장학금을 미수혜 했을 경우도 약 1.10배 위험도가 높았고, 특히 이체일자는 26일 이상 즉, 월말보다 월초일 수록 미회수 위험도가 높아졌다.

Table 3.2 Result of logistic regression analysis

Model 1			Model 2			Model 3		
Variable	β	OR	Variable	β	OR	Variable	β	OR
Intercept	-0.3375		Intercept	0.0295		Intercept	-0.8219**	
Age(ref: ≤ 20)			Sex(ref: Female)			Sex(ref: Female)		
21-23	-0.0191	0.98	Male	0.1292**	1.14	Male	-0.0554	0.95
24-26	0.2019	1.22	Age(ref: ≤ 20)			Age(ref: ≤ 20)		
27-30	-0.2558	0.77	21-23	0.0151	1.02	21-23	0.0363	1.04
≥ 31	-0.0225	0.98	24-26	-0.0169	0.98	24-26	0.133*	1.14
Household income(ref: Class 0)			27-30	0.0172	1.02	27-30	0.1769*	1.19
Class 1	-0.2868	0.75	≥ 31	0.083	1.09	≥ 31	0.2773**	1.32
Class 2-4	-0.3675	0.69	Household income(ref: Class 0)			Marital status(ref: Not married)		
Class 5-6	-0.6808	0.51	Class 1	-0.1202	0.89	Married	0.0499	1.05
Class 7-9	-0.6119	0.54	Class 2-4	-0.152	0.86	Household income(ref: Class 0)		
Class 10	-0.9327**	0.39	Class 5-7	-0.2871**	0.75	Class 1	-0.1794	0.84
etc	-0.7525*	0.47	Class 8-9	-0.3224**	0.72	Class 2-4	-0.1025	0.90
Level of schooling(ref: Undergraduate)			Class 10	-0.3387**	0.71	Class 5-7	-0.1292	0.88
Graduate	0.3235	1.38	etc	-0.2495**	0.78	Class 8-9	-0.1882*	0.83
Overdue account(ref: ≥ 10)			Field of study(ref: Humanities)			Class 10	-0.1971	0.82
1	1.248*	3.48	Social Sciences	-0.0234	0.98	etc	-0.1519	0.86
2	0.7253	2.07	Education	0.0301	1.03	Level of schooling(ref: Graduate)		
3-4	0.5816	1.79	Engineering	-0.0398	0.96	Undergraduate	0.1725*	1.19
5	0.1458	1.16	Natural Sciences	-0.0888	0.92	Field of study(ref: Humanities)		
6-9	0.3099	1.36	Medical Science	0.0954	1.10	Social Sciences	-0.0652	0.94
Monthly repayment(ref: < 28000)			Arts & Physical	-0.0623	0.94	Education	-0.0452	0.96
28,000-50,000	-0.1587	0.85	Overdue account(ref: ≥ 10)			Engineering	-0.1733**	0.84
50,000-65,000	0.061	1.06	1	0.6232**	1.88	Natural Sciences	-0.1092	0.90
65,000-80,000	-0.2452	0.78	2	0.346**	1.41	Medical Science	-0.2065*	0.81
80,000-110,000	0.0366	1.04	3-7	0.1506	1.16	Arts & Physical	-0.1429*	0.87
110,000-250,000	-0.00195	1.00	8-9	0.1166	1.12	Overdue account(ref: ≥ 10)		
$\geq 250,000$	-0.1357	0.87	Monthly repayment(ref: $< 28,000$)			1	1.2684**	3.56
Balance(ref: $\geq 13,000,000$)			28,000-50,000	0.0107	1.01	2	0.8984**	2.46
$< 1,200,000$	0.5052	1.66	50,000-65,000	0.1494*	1.16	3-7	0.5522**	1.74
1,200,000-2,200,000	0.4577	1.58	65,000-80,000	0.0569	1.06	8-9	0.307	1.36
2,200,000-3,400,000	0.0295	1.03	80,000-110,000	0.1702*	1.19	Monthly repayment(ref: $< 28,000$)		
3,400,000-6,000,000	-0.2456	0.78	110,000-250,000	0.1221	1.13	28,000-50,000	0.0316	1.03
6,000,000-13,000,000	-0.5503*	0.58	$\geq 250,000$	0.0394	1.04	50,000-65,000	0.0692	1.07
Arrears over balance(ref: 3.3-15.8)			Arrears(ref: $< 15,000$)			65,000-80,000	0.0814	1.09
< 0.4 or ≥ 15.8	0.4185*	1.52	15,000-30,000	-0.2388**	0.79	80,000-110,000	0.1339	1.14
0.4-0.8	0.0547	1.06	30,000-60,000	-0.4165**	0.66	110,000-250,000	0.218**	1.24
0.8-1.6	0.2398	1.27	60,000-130,000	-0.3957**	0.67	$\geq 250,000$	0.5328**	1.70
1.6-3.3	-0.1021	0.90	130,000-260,000	-0.5007**	0.61	Arrears(ref: $< 45,000$)		
Overdue record(ref: No)			260,000-680,000	-0.3558**	0.70	45,000-100,000	-0.2744**	0.76
Yes	0.1824	1.20	$\geq 680,000$	0.1924	1.21	100,000-270,000	-0.1053	0.90
Day in arrears(ref: ≤ 10)			Balance(ref: $\geq 7,500,000$)			270,000-500,000	-0.0692	0.93
11-28	0.2537	1.29	$< 1,000,000$	0.9171**	2.50	500,000-960,000	0.0664	1.07
29	0.8107**	2.25	1,000,000-1,600,000	0.7242**	2.06	960,000-1,700,000	0.3273**	1.39
≥ 30	0.1262	1.13	1,600,000-2,800,000	0.5652**	1.76	$\geq 1,700,000$	0.6833**	1.98
Loan interest rate(ref: 2.7)			2,800,000-4,000,000	0.2341**	1.26	Balance(ref: $\geq 12,000,000$)		
2.9, 3.9	0.1189	1.13	4,000,000-7,500,000	0.0582	1.06	$< 1,000,000$	0.7966**	2.22
4.9, 5.2	0.0133	1.01	Grace period(ref: ≥ 73 months)			1,000,000-1,600,000	0.3356**	1.40
5.7	-0.1887	0.83	≤ 12	0.0454	1.05	1,600,000-2,800,000	0.2779**	1.32
5.8	0.3932	1.48	13-36	0.1963*	1.22	2,800,000-3,300,000	0.2805*	1.32
Due date for payment(ref: ≥ 26)			37-48	0.0366	1.04	3,300,000-12,000,000	0.2185*	1.24
≤ 5	1.2096**	3.35	49-60	0.1184	1.13	Overdue record(ref: No)		
6-10	0.9201**	2.51	61-72	0.1202	1.13	Yes	0.1676**	1.18
11-15	0.6047*	1.83	Overdue record(ref: No)			Base year of loan(ref: 2015)		
16-20	1.0273**	2.79	Yes	0.1414**	1.15	2009	0.4145**	1.51
21-25	0.3458	1.41	Day in arrears(ref: ≤ 35)			2010	0.2495**	1.28
			36-40	0.1868**	1.21	2011	0.1987*	1.22
			41-45	0.3922**	1.48	2012	0.1782*	1.20
			46-50	0.8065**	2.24	2013	0.1205	1.13
			≥ 51	1.0208**	2.78	2014	0.1338	1.14
			National grant(ref: Yex)			Due date for payment(ref: ≥ 26)		
			No	0.093**	1.10	≤ 5	1.3768**	3.96
			Base year of loan(ref: 2015)			6-10	1.0137**	2.76
			2009	0.3499**	1.42	11-15	0.8089**	2.25
			2010	0.1906**	1.21	16-20	0.4305**	1.54
			2011	0.1461*	1.16	21-25	0.21**	1.23
			2012	0.0499	1.05	Number of overdue in recent 3 months(ref: 1)		
			2013	0.0732	1.08	2	0.7162**	2.05
			2014	0.0157	1.02	3	2.0255**	7.58
			Due date for payment(ref: ≥ 26)					
			≤ 5	0.8754**	2.40			
			6-10	0.6975**	2.01			
			11-15	0.5978**	1.82			
			16-20	0.4453**	1.56			
			21-25	0.2627**	1.30			

* $p < 0.05$, ** $p < 0.01$

연체 3개월 이상인 Model 3에서는 연령이 높아질수록 미회수 연체 위험이 높았으며, 학부보다 대학원에서 미회수 위험이 약 1.19배 높았고, 인문계열보다 공학, 의약, 예체능계열의 위험도가 낮았다. 연체계좌 수가 10개 이상인 경우보다 1개는 약 3.56배, 2개는 약 2.46배, 3~7개는 약 1.74배 미회수 위험도가 높았으며, 월평균 상환액이 28,000원 미만인 경우보다 110,000~250,000원인 경우가 약 1.24배, 250,000원 이상은 약 1.70배 미회수 위험이 높았다. 연체액이 약 45,000원 미만인 경우보다 45,000~100,000원인 경우는 미회수 위험이 낮았지만, 960,000~1,700,000원은 약 1.39배, 1,700,000원 이상은 약 1.98배 미회수 위험이 높았다. 연체잔액을 살펴보면, 12,000,000원 이상인 경우보다 연체금액이 낮아질수록 미회수 위험도가 높았으며, 이체일자 역시 Model 2와 같이 월말보다 월초일 수록 미회수 위험도가 높아졌다. 특히 Model 3의 최근 3개월 이내 연체 횟수가 2번인 경우는 약 2.05배, 3번은 약 7.58배 미회수 위험도가 높은 결과를 보였다.

3.2. 신용 위험 평점표

앞서 개발된 추정된 로지스틱 회귀모형을 이용하여 적합한 신용위험 평점표를 만들고, 이 과정에서 다양한 방법을 고려할 수 있지만 여기서는 전형적인 방법 중 하나인 PDO (points to double the odds) 즉, 오즈를 두 배로 만드는 지점의 수치를 이용해 평점표를 작성하는 방식으로 진행하였다.

먼저 Model별 각 변수들에 대해 추정된 회귀계수에서 가장 작은 회귀계수 추정치를 빼서 회귀계수 추정치를 0보다 크거나 같도록하여 보정된 회귀계수를 계산한다. 이후 적당한 PDO를 정하여 식 (3.1) 와 같이 보정된 회귀계수를 하나의 점수로 선형 변환시킨다 (Kang 등, 2014).

$$Score = adjusted \hat{\beta} \times \left[\frac{PDO}{\log(2)} \right] \tag{3.1}$$

본 연구에서는 PDO를 50으로 정했으며, Model별 평점표 결과는 Table 3.3과 같다.

Table 3.3 Result of scorecard

Model 1		Model 2		Model 3	
Variable	Score	Variable	Score	Variable	Score
Age		Sex		Sex	
≤ 20	18.5	Female	0.0	Female	4.0
21-23	17.1	Male	9.3	Male	0.0
24-26	33.0	Age		Age	
27-30	0.0	≤ 20	1.2	≤ 20	0.0
≥ 31	16.8	21-23	2.3	21-23	2.6
Household income		24-26	0.0	24-26	9.6
Class 0	67.3	27-30	2.5	27-30	12.8
Class 1	46.6	≥ 31	7.2	≥ 31	20.0
Class 2-4	40.8	Household income		Marital status	
Class 5-6	18.2	Class 0	24.4	Not married	0.0
Class 7-9	22.6	Class 1	15.8	Married	3.6
Class 10	0.0	Class 2-4	13.5	Household income	
etc	13.0	Class 5-7	3.7	Class 0	14.2
Level of schooling		Class 8-9	1.2	Class 1	1.3
Undergraduate	0.0	Class 10	0.0	Class 2-4	6.8
Graduate	23.3	etc	6.4	Class 5-7	4.9
Overdue account		Field of study		Class 8-9	0.6
1	90.0	Humanities	6.4	Class 10	0.0
2	52.3	Social Sciences	4.7	etc	3.3
3-4	42.0	Education	8.6	Level of schooling	
5	10.5	Engineering	3.5	Graduate	0.0
6-9	22.4	Natural Sciences	0.0	Undergraduate	12.4
> 10	0.0	Medical Science	13.3	Field of study	
Monthly repayment		Arts & Physical	1.9	Humanities	14.9
< 28,000	17.7	Overdue account		Social Sciences	10.2
28,000-50,000	6.2	1	45.6	Education	11.6
50,000-65,000	22.1	2	25.0	Engineering	2.4
65,000-80,000	0.0	3-7	10.9	Natural Sciences	7.0
80,000-110,000	20.3	8-9	8.4	Medical Science	0.0
110,000-250,000	17.5	≥ 10	0.0	Arts & Physical	4.6
≥ 250,000	7.9	Monthly repayment		Overdue account	
Balance		< 28,000	0.0	1	91.5
< 1,200,000	76.1	28,000-50,000	0.8	2	64.8
1,200,000-2,200,000	72.7	50,000-65,000	10.8	3-7	39.8
2,200,000-3,400,000	41.8	65,000-80,000	4.1	8-9	22.1
3,400,000-6,000,000	22.0	80,000-110,000	12.3	≥ 10	0.0
6,000,000-13,000,000	0.0	110,000-250,000	8.8	Monthly repayment	
≥ 13,000,000	39.7	≥ 250,000	2.8	< 28,000	0.0
Arrears over balance		Arrears		28,000-50,000	2.3
<0.4 or ≥ 15.8	37.6	< 15,000	36.1	50,000-65,000	5.0
0.4-0.8	11.3	15,000-30,000	18.9	65,000-80,000	5.9
0.8-1.6	24.7	30,000-60,000	6.1	80,000-110,000	9.7
1.6-3.3	0.0	60,000-130,000	7.6	110,000-250,000	15.7
3.3-15.8	7.4	130,000-260,000	0.0	≥ 250,000	38.4
Overdue record		260,000-680,000	10.5	Arrears	
No	0.0	≥ 680,000	50.0	< 45,000	19.8
Yes	13.2	Balance		45,000-100,000	0.0
Day in arrears		< 1,000,000	66.2	100,000-270,000	12.2
≤ 10	0.0	1,000,000-1,600,000	52.2	270,000-500,000	14.8
11-28	18.3	1,600,000-2,800,000	40.8	500,000-960,000	24.6
29	58.5	2,800,000-4,000,000	16.9	960,000-1,700,000	43.4
≥ 30	9.1	4,000,000-7,500,000	4.2	≥ 1,700,000	69.1
Loan interest rate		≥ 7,500,000	0.0	Balance	
2.7	13.6	Grace period		< 1,000,000	57.5
2.9, 3.9	22.2	≤ 12	3.3	1,000,000-1,600,000	24.2
4.9, 5.2	14.6	13-36	14.2	1,600,000-2,800,000	20.0
5.7	0.0	37-48	2.6	2,800,000-3,300,000	20.2
5.8	42.0	49-60	8.5	3,300,000-12,000,000	15.8
Due date for payment		61-72	8.7	≥ 12,000,000	0.0
≤ 5	87.3	≥ 73 months	0.0	Overdue record	
6-10	66.4	Overdue record		No	0.0
11-15	43.6	No	0.0	Yes	12.1
16-20	74.1	Yes	10.2	Base year of loan	
21-25	24.9	Day in arrears		2009	29.9
≥ 26	0.0	≤ 35	0.0	2010	18.0
		36-40	13.5	2011	14.3
		41-45	28.3	2012	12.9
		46-50	58.2	2013	8.7
		≥ 51	73.6	2014	9.7
		National grant		2015	0.0
		Yes	0.0	Due date for payment	
		No	6.7	≤ 5	99.3
		Base year of loan		6-10	73.1
		2009	25.2	11-15	58.3
		2010	13.7	16-20	31.1
		2011	10.5	21-25	15.1
		2012	3.6	≥ 26	0.0
		2013	5.3	Number of overdue in recent 3 months	
		2014	1.1	1	0.0
		2015	0.0	2	51.7
		Due date for payment		3	146.1
		≤ 5	63.1		
		6-10	50.3		
		11-15	43.1		
		16-20	32.1		
		21-25	18.9		
		≥ 26	0.0		

평점표에서 영향력을 나타내는 범위를 Model별로 살펴보면, Model 1에서는 과거연체경험이 최대평점 13.2로 가장 작고, 연체계좌 수의 최대평점이 90.0으로 가장 큰 것을 확인할 수 있었으며, Model 2에서는 국가장학금 수혜유무가 최대평점 6.7로 가장 작고, 연체일수의 최대평점이 73.6으로 가장 큰 것을 알 수 있다. 마지막으로 Model 3에서는 결혼 유무의 최대평점이 3.6으로 가장 작고, 최근 3개월 이내 연체횟수의 최대평점이 146.1로 가장 큰 것을 확인할 수 있었다. 예를들어, 연체 1개월 모델 즉, Model 1에 해당되는 연체자가 연령이 24~26세, 소득분위가 0분위, 대학원생, 연체계좌가 1개, 월평균 상환액이 50,000~65,000원, 연체잔액이 1,200,000원 미만, 연체잔액 대비 연체액이 0.4미만이거나 15.8이상이고, 과거 연체경험이 있으며, 연체일수가 29일, 대출금리 5.8%를 이용하고, 이체일자가 5일 이전으로 된 연체자는 $33.0+67.3+23.3+90.0+22.1+76.1+37.6+13.2+58.5+42.0+87.3=550.4$ 점이 된다. Model 1에서 11개의 항목에 해당되는 평점이 550.4점으로 가장 높은 미회수 위험이 높은 점수를 갖는 연체자로 판단할 수 있다.

4. 결론

본 연구에서는 학국장학재단의 일반상환 학자금 대출 연체자를 대상으로 로지스틱 회귀분석을 이용하여 연체 회수 모형을 연체기간에 따라 3가지 Model로 개발하고 그에 따라 신용 평점표를 작성하였다.

앞서 결과를 통해 연체채권은 시간이 경과함에 따라 채권의 속성이 바뀐다는 것을 확인할 수 있었다. 먼저 연체 1개월 모형 (Model 1)에서는 연체계좌수가 적은 연체자, 이체일자가 5일 이전인 연체자, 연체잔액이 2,200,000원 미만인 연체자, 소득분위가 0분인 연체자일수록 미회수 평점이 높게 나타났으나, 연체 2개월 모형 (Model 2)에서는 연체 일수가 51일 이상인 연체자, 연체잔액이 1,000,000원 미만인 연체자, 이체일자가 5일 이전인 연체자, 연체금액이 680,000원 이상인 연체자 일수록 미회수 평점이 높게 나타났다. 마지막으로 연체 3개월 이상 모형 (Model 3)에서는 최근 3개월 이내 연체 횟수가 3회인 연체자, 이체일자가 5일 이전인 연체자, 연체계좌수가 적은 연체자, 연체액이 1,700,000원 이상인 연체자 일수록 미회수 평점이 높게 나타났다.

현행 학자금대출은 원칙적으로 ‘요건대출’로 일정 요건을 만족시키는 경우 별도의 신용평가 없이 미래의 소득 발생을 전제로 대출이 실행되는 성격을 지니고 있어 담보대출이나 일정한 소득을 가진 대출자를 대상으로 하는 일반적인 신용 대출에 비하여 신용위험이 높은 편이며, 정책금융적 성격으로 인해 채권관리에 있어서도 적극적 관리에 어려움이 존재하고 있다.

본 연구에서 개발된 연체 회수 모형이나 신용 평점표를 바탕으로 연체 채권관리를 함에 있어 좀더 세분화된 관리서비스를 제공하고, 상담센터의 상담원이 연체자의 평점에 따라 상담전략을 세우는데 기초 자료가 될 수 있을 것으로 사료된다.

다만, 연구에 활용된 분석 자료 중 소득분위에서 소득이 파악되지 않는 ‘기타’의 정의나 연체자의 금융정보 등을 추가적으로 활용한 모형 개발을 고려할 필요가 있다. 또한 대출자의 취업현황이나 졸업후 상황을 확인할 수 있는 변수를 찾아내어 고려하고, 본 연구에서는 금융기관의 선례에 따라 모형별 연체 기간을 1개월 단위로 구분하였지만 향후 학자금대출 상황에 맞게 조정하는 것이 모형의 고도화에 필요한 요건으로 보인다. 마지막으로 본 연구에서는 로지스틱 회귀분석 기법만을 적용하였으나, 추후 타 모형 (의사결정나무, 신경망분석 등)과의 성능비교에 대한 추가적인 연구를 진행하고자 한다.

References

- Altman, E. I., Haldeman, R. and Narayanan, P. (1977). Zeta analysis-a new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance*, 1, 29-54.

- Choi, J. S., Han, J. T., Kim, M. J. and Jeong, J. (2015). Developing the high risk group predictive model for student direct loan default using data mining. *Journal of the Korean Data & Information Science Society*, **26**, 1417-1426.
- Grunert, J. and Weber, M. (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking and Finance*, **33**, 505-513.
- Jin, S. K., Kim, K. R. and Park, C. (2012). Cutpoint selection via penalization in credit scoring. *The Korean Journal of Applied Statistics*, **25**, 261-267.
- Khieu, H. D., Mullineaux, D. J. and Yi, H. C. (2012). The determinants of bank loan recovery rates. *Journal of Banking and Finance*, **36**, 923-933.
- Kang, H. C., Han, S. T., Choi, J. H., Lee, S. G., Kim, E. S. and Um, I. H. (2014). *Methodology of data mining for big data analysis: A case study on SAS Enterprise Miner*, Free Academy, Seoul.
- Kim, T. H. and Kim, Y. H. (2013). A study on the analysis of customer loan for the credit finance company using classification model. *Journal of the Korean Data & Information Science Society*, **24**, 411-425.
- Leow, M. and Crook, J. (2016). The stability of survival model parameter estimates for predicting the probability of default: Empirical evidence over the credit crisis. *European Journal of Operational Research*, **249**, 457-464.
- Pompe, P. P. M. and Bilderbeek, J. (2005). The prediction of bankruptcy of small-and medium-sized industrial firms. *Journal of Business Venturing*, **20**, 847-868.
- Siddiqi, N. (2006). *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, John Wiley & Sons, New Jersey.
- Woo, H. S., Lee, S. H. and Cho, H. J. (2013). Building credit scoring models with various types of target variables. *Journal of the Korean Data & Information Science Society*, **24**, 85-94.
- Zurada, J. and Zurada, M. (2002). How secure are good loans: Validating loan-granting decisions and predicting default rates on consumer loans. *The Review of Business Information Systems*, **6**, 65-83.

Developing the credit risk scoring model for overdue student direct loan

Jun-Tae Han¹ · Jina Jeong²

^{1,2}Statistics & Analysis Team, Korea Student Aid Foundation

Received 26 July 2016, revised 9 September 2016, accepted 19 September 2016

Abstract

In this paper, we develop debt collection predictive models for the person in arrears by utilizing the direct loan data of the Korea Student Aid Foundation. We suggest credit risk scorecards for overdue student direct loan using the developed 3 models. Model 1 is designed for 1 month overdue, Model 2 is designed for 2 months overdue, and Model 3 is designed for overdue over 2 months. Model 1 shows that the major influencing factors for the delinquency are overdue account, due data for payment, balance, household income. Model 2 shows that the major influencing factors for delinquency loan are days in arrears, balance, due date for payment, arrears. Model 3 shows that the major influencing factors for delinquency are the number of overdue in recent 3 months, due data for payment, overdue account, arrears. The debt collection predictive models and credit risk scorecards in this study will be the basis for segmented management service and the call & collection strategies for preventing delinquency.

Keywords: Credit risk scorecard, logistic regression, student direct loan.

¹ Corresponding author: Team Manager, Statistics & Analysis Team, Korea Student Aid Foundation, 125 Sinam-ro, dong-gu, Daegu 41200, Korea. E-mail: hanjt@kosaf.go.kr

² Staff member, Statistics & Analysis Team, Korea Student Aid Foundation, 125 Sinam-ro, dong-gu, Daegu 41200, Korea.