머신러닝 알고리즘을 이용한 동태적 자본 구조 예측

이정환*, 조진형**, 김봉준*** 이원응****

초 록

본 연구는 머신러닝(ML) 알고리즘을 활용하여 기업의 목표 부채비율과 이를 예측하는데 필요한 결정 변수를 살펴보고자 한다. 구체적으로 2003~2019년의 기업 표본 가운데 2003~2014년을 표본 내 데이터, 2015~2019년을 표본 외 데이터로 정했으며, 선형 모형인 다중회귀, 라쏘와 머신러닝 모형인 랜덤포레스트, GBM(Gradient Boosting Regression)을 비교 분석하였다. 부채비율에 대하여 예측한 결과 랜덤포레스트와 GBM은 다중회귀와 라쏘에 비하여 결정계수(R^2_{OS})가 높고 평균 제곱 오차가 낮은 것으로 나타나 높은 설명력을 확인할 수 있었다. 특히 랜덤포레스트 모형의 경우 결정 요인은 시장가-장부가 비율, 순배당율, 알트만 Z-점수, 수익성 순으로 높은 것으로 분석되었다. 이어 기업 목표 부채비율로의 조정속도를 추정한 결과 머신러닝 모형인 랜덤포레스트와 GBM이, 선형 모형인 라쏘와 다중회귀 모형에 비해 50% 빠르다는 것을 확인하였다.

키워드 : 머신러닝, 부채비율, 부채비율 조정 속도, 최적 자본 구조

* 제1저자, 한양대 경제금융대학 부교수, jeonglee@hanyang.ac.kr

** 교신저자, 한양대 경제금융대학 박사과정, enish27@hanyang.ac.kr

*** 공저자, 한양대 경제금융대학 박사과정, kbj0918@hanyang.ac.kr

**** 공저자, 한양대 경제금융대학, woneunglee@hanyang.ac.kr

Ⅰ. 서론

기업의 부채비율(Leverage)[1]은 재정 상태, 혹은 재정 건전성을 나타내는 대표적인 기업 경영 지표이다. 부채비율은 기업의 부채총액이 총자산, 혹은 자기자본과 비교하여 얼마나 되는지 확인하는데 쓸 수 있다. 부채비율이 높다면 기업의 재정 상태가 나쁘거나 최악엔 파산에 이를 수 있으며, 부채비율이 낮다면 기업의 차입에 대한 의존도가 적다는 것을 알 수 있다. 취약한 재무구조와 더불어 과다한 부채비율은 기업의 구조조정으로 이어질 수 있는데, 한국의 경우 IMF 외환위기를 이후인 1999년 말까지 부채비율 200% 달성이라는 목표 부채비율이 대기업을 중심으로 요구된 바 있다. 이를 계기로 우리나라는 대기업을 중심으로 부채비율 축소가 주요 재무 전략으로 자리잡았다(이원흠 외, 2001; 박연희 외, 2011).

기업의 부채비율 조정과 관련해선 여러 이론이 존재한다. 먼저 상충이론(trade-off theory)은 기업의 부채가 이 이익과 비용 간의 상쇄로 인하여 정해진다는 사실을 주목하고 있다. 부채를 쓰면 외부투자자와 내부 경영자 간 대리인 문제가 생기는 한편, 부채는 주주-채권자 간의 대리인 문제를 해결해줄 수 있다는 것이다(Jensen and Meckling, 1976; Jensen, 1986). 이에 Stulz(1990)와 Morellec(2004)은 부채가 주주-채권자 간의 이해상충 문제를 악화시킨다고 주장한다. 반면 Myers(1984)가 주장한 순서 이론은 기업이 유보이익과 부채를 통해 우선적으로 자본 조달하고 마지막으로 주식시장을 이용한다는 이론이다. 이는 경영자와 투자자 간에 존재하는 정보의 비대칭성에서 초래된 것이라는 것이 이론의 핵심이다. 또 Baker와 Wurgler(2002)이 제시한 시장 타이밍 이론은, 주식시장이 호황일 때 주식발행에 의한 자본조달이 증가하지만, 불황이라면 부채 발행에 의한 자본조달이 증가한다는 사실을 언급하고 있다. 기업의 주식이 과대평가 되어있을 때 자본구조에 변화가 온다는 것이다.

기업의 자본구조와 부채비율에 대한 이론적 근거는 다양하지만, 기존 연구는 기업이 도달하고자 하는 목표 부채비율(Target Leverage)은 다양한 변수를 도입하여 자본구조 결정 요인을 검토하고, 이들 요인을 활용하여 목표 부채비율로의 조정 속도를 추정하고 있다. 목표 부채비율로의 조정 속도 추정은 기업들이 상충이론에서 예측한 목표 부채비율로 얼마나 빠른 속도로 수렴하고 있는지를 보여줌으로써, 기업의 최적 자본구조에 대해 상충이론이 얼마나 설명력을 갖추고 있는지를 확인할 수 있는 중요한

[1] 부채비율은 부채비중을 나타내는 대표적인 측정치다. 부채비율, 자본구조, 레버리지(Leverage)는 모두 같은 의미이다. 본 논문에서는 '부채비율'로 통일해 쓰고자 한다.

방법으로 꼽힌다. 만약 부채비율 조정속도가 1로 추정된다면 현실 기업의 자본구조가 절충이론에 완전히 일치한다는 것을 의미하며, 0으로 추정된다면 상충이론이 현실 기업에 대해 전혀 설명력을 갖지 못한다는 것을 의미한다. Graham et al. (2001)은 기업 대상 설문 조사를 언급하며 목표 부채비율이 기업의 자본구조 결정을 내리는데 중요한 고려사항이라는 사실을 언급한 바 있다. 최윤이(2015)는 실질적으로 부채비율을 감소시키는 방법으로 수익성이 낮은 자산을 매각하여 타인자본의 상환, 유상증자를 통해 조달된 자금으로 타인자본을 실질적으로 감소시켜서 자기자본을 증가시키거나, 타인자본을 자기자본으로 전환해서 감소시키는 방법 등을 언급하고 있다. 이러한 방법들 활용해 수익성을 개선하거나 재무건전성을 높일 수 있으며 차입금 의존도도 낮출 수 있다는 것이다.

국내외 연구는 고정효과 패널모형, Bruno(2005), OLS, FM, GMM 등 다양한 전통적 분석법을 활용하여 최적 부채비율과 이에 대한 조정 속도를 예측하였다(Aybar-Aria et al., 2012; Fama et al., 2002; 김영래 외, 2007; 윤보현 외, 2016; 이원흠 외, 2001). 일부 선행연구는 이와 같이 전통적인 계량분석을 활용할 경우 발생할 수 있는 한계를 언급하고 있다. 윤보현 외(2016)는 부채비율 조정속도 추정 과정에서 기업 자료의 특수성, 즉 패널 구조의 불균형, 종속 변수들의 절단, 횡단면 데이터의 수량부족 등의 문제들이 추정치에 편의를 발생시킨다는 사실을 지적하고 있다. Amini et al. (2021) 역시 전통적인 계량 분석을 활용할 경우 비선형성과 복잡한 상호관계가 무시될 경우 상관성이 높은 부채비율 결정 요인을 찾는 것이 어렵다는 사실을 지적하고 있다 (Frank et al, 2009). 이들은 머신러닝(ML)과 같은 방법론을 활용한다면 데이터 차원을 축소시키거나, 상관성이 높은 중복 변수의 분산(redundant variation)을 줄임으로써 많은 결정 변수를 정확하게 분류할 수 있다고 주장하였다.

이를 감안하여 본 연구는 머신러닝 기법을 활용하여 국내 상장 기업의 목표 부채비율을 예측하고, 목표 부채비율로의 조정 속도를 분석한다. 특히 2003년부터 2019년까지 한국의 코스피(KOSPI) 상장 이력이 있는 기업을 대상으로 재무와 거시경제 지표를 입수하여 국내 상장기업의 자본구조 결정 요인을 검토하고, 머신러닝 모델을 활용하여 목표 부채 비율로의 조정 속도를 추정하고자 한다. 구체적으로 선형회귀(LM), 라쏘(LASSO), 랜덤포레스트(Random Forest, RF), GBM(Gradient Boosting Regression) 모델을 활용하여 모델별 성능을 비교 분석하고자 한다.

본 연구의 주요 분석 결과는 다음과 같다. 첫째, 기업의 목표 부채비율을 예측 시 랜덤포레스트와 GBM은 다중회귀와 라쏘에 비해 결정계수(R^2_{OS})가 높고 평균 제곱 오차

(Mean Squared Error)가 낮은 것으로 나타났다. 둘째, 랜덤 포레스트와 GBM 모델을 통해 변수 중요성(Variable Importance)을 분석한 결과 시장가-장부가 비율(Market-to-Book ratio), NeyPay(순발행액), Z-Score, Profit(매출액), Tangibility(유형자산) 순인 것으로 나타났다. 셋째, 각 모델별로 목표 부채비율로의 부채비율 조정 속도(Speed of Adjustment)를 분석한 결과, 랜덤포레스트, GBM, 다중회귀, 라쏘 순으로 부채비율 조정 속도가 빠른 것으로 나타난다. 이와 같은 연구 결과는 기업의 최적 자본구조와 목표 부채비율에 대한 부채비율 조정 속도를 머신러닝 기법으로 예측한 최근 연구 결과(Amini et al., 2021)와 대체적으로 일치한 것으로 나타난다.

본 연구의 대표적인 공헌점은 다음과 같다. 첫째, 국내 코스피 시장에 상장된 이력이 있는 기업 전체를 대상으로 최적의 자본구조를 예측하는 머신러닝 분석을 실시하였다는 것이다. 이와 관련된 기존 연구(박연희 외, 2011; 윤보현 외, 2016; 이원흠 외, 2001)는 상당수 있었지만, 선형 모형과 머신러닝 모형을 활용하여 두 모형 간 성능을 비교 분석한 부채비율 관련 연구는 본 연구가 국내에선 최초인 것으로 보인다. 둘째, 다양한 재무·거시경제 환경에서의 머신러닝 분석 기법의 이점을 증명하였다는 것이다. 특히 머신러닝 모형이 과적합성, 내생성과 관련된 여러 연구에 폭넓게 쓰일 수 있을 것으로 기대된다. Ⅱ장에서는 기업의 자본구조와 부채비율에 대한 선행연구를 살펴보고, 이어 Ⅲ장에서는 연구 방법론을 소개한다. Ⅳ장에서는 연구 결과를 제시하고, Ⅴ장은 결론을 내리고자 한다.

Ⅱ. 이론적 배경

기업 부채 비율에 관한 국내 연구들은 주로 기업들의 자본구조 결정요인에 관한 연구 및 기업들의 행동패턴들이 절충이론과 순서이론에 잘 부합하는지에 관한 연구들이 주종을 이루고 있으나, 최근에는 자본구조 조정속도 추정에 관한 연구도 비교적 활발하게 진행되고 있다. 부채비율 조정속도 추정에 관한 대표적인 연구들은 재무적 제약이 부채비율 조정속도에 미치는 영향을 분석한 신민식·김수은(2008), 재벌과 비재벌, 대기업과 소기업 집단으로 구분하여 부채비율 조정 속도를 차이를 비교한 손판도·손승태(2008), 자본조달 시장 접근성에 따른 부채비율 조정 속도의 차이에 관해 연구한 김진수(2010), 기업의 현금흐름이 부채비율 조정 속도에 미치는 영향을 추정한 신민식·김수은(2012) 등이 있다.

기업 경영자는 목표 부채비율(target leverage)로 자본구조를 변경함으로써 최적의 자본구조(Optimal capital structure)를 추구할 유인을 가지고 있다(Flannery and Rangan, 2006; Frank and Goyal, 2004; Graham and Harvey, 2001; 윤보현 외, 2016; 윤봉한, 2005; 이원흠 외, 2001). 구체적으로 기업 경영자는 목표 부채 비율을 추구함으로써 당기, 혹은 차기의 자본 조달에 있어 재무적인 유연성을 유지하고자 한다. 그러나 현실적으로 기업의 실제 부채비율은 목표 부채비율과 차이가 발생하며, 기업은 최적 자본구조로부터 일시적으로 이탈할 수 있다(손승태 외, 2007). 궁극적으로 기존 문헌은 최적 부채비율로의 조정 속도(비용)을 언급하고 있다. 자본구조의 기업 부채비율은 최적 자본구조와 차이가 있는데, 기업의 부채비율 조정과정에서 발생하는 조정비용(속도)은 기업이 신속하게 부채비율을 목표 부채비율로 조정하는 것을 지연시키는 경향이 있다(Shyam-Sunder and Myers, 1999; Miguel and Pindado, 2001).

기업의 부채비율을 다룬 대표 이론은 상충 이론(trade-off theory)이다. 이 이론은 기업이 부채로부터 얻는 이득과 부채로 인해 발생하는 비용 간의 상충 관계(trade-off)가 존재한다는 사실을 제시하고 있다. 이 이론에 의하면 기업엔 기업의 가치를 최대화시키는 최적의 부채 비율이 존재한다. 구체적으로 기업이 부채를 이용한다면 이자 비용의 절세 효과로 인하여 기업 가치가 증가되며, 반면 부채비율이 높아져 파산 비용 증가하면 앞서 절세를 통한 기업가치 증가분이 상쇄될 수 있다. 만약 기업의 부채비율이 목표 비율로부터 이탈한다면 기업은 목표 비율을 향해 당기의 부채비율을 조정할 수 있다(김영래 외, 2007). 이런 측면에서 기업이 최적 부채비율에서 부채를 쓰는 건 기업가치 측면에서 합리적인 의사결정이라고 보는 추세라고 할 수 있다.

반면 Myers et al.(1984)가 제시한 자본조달 순위 이론(Pecking order theory)은 정보 비대칭과 기업 자본 비용의 차이로 인하여 기업은 이익유보금을 활용한 내부 금융으로 자본을 조달하며, 추후 외부금융이 필요하면 먼저 부채로 자본을 조달하며, 마지막으로 자기자본으로 자본을 조달한다는 특징을 언급하고 있다. 결국 경영자와 외부 투자자(주주) 간 정보비대칭에 따른 자금 조달 우선 순위의 인위적인 특성으로 인하여 최적의 자본구조가 부정된다는 것이다. 끝으로 Baker et al.(2002)에 의하여 강조된 시장 타이밍 이론(Market timing theory)은 주식 시장이 호황이면 주식 발행을 통해 자본을 조달하고, 주식시장이 불황이면 부채 발행을 통해 자본을 조달한다. 이 경우 기업의 주식이 과대평가 되어있을 때 기업의 자본구조가 변화됨에 따라 최

적 자본구조의 존재가 왜곡될 수 있다(김영래 외, 2007).

기업의 부채비율에 대한 최근 연구는 재무 변수 뿐 아니라 시장 변수와 거시경제 지표까지 아울러 부채비율 결정 요인을 제시하고 있다. 특히 인플레이션, 국내총생산 성장률, 국채 10년물과 1년물의 수익률 차이(김영래 외, 2007; Amini et al, 2021; Frank et al.,2004) 등을 주요 경제 변수로 활용하고 있다. 이상의 선행연구를 고려하여 본 연구는 다양한 재무·거시경제 변수를 활용하여 목표 부채비율을 예측하는데 중요한 영향을 미치는 요인을 살펴볼 것이다.

Ⅲ. 방법론

Ⅲ-1. 분석 표본

본 연구의 분석 표본은 2003년부터 2020년까지 KOSPI 상장 이력이 있는 기업 중 비금융·보험 기업이 대상이다. 일부 연구에서는 신규 상장 기업 혹은 상장폐지 기업을 누락시킴으로써 생존편의(Survivorship bias) 문제를 발생시키고 있다. 본 연구는 이러한 문제를 해소하고자 분석 기간 중 신규 상장기업과 상장폐지 기업을 모두 분석 표본에 포함시켰다. 각 종속 변수와 설명 변수의 극단 값을 1% 수준으로 제거한 결과 기업 숫자는 686개, 관측치는 6,545개에 이른다. 관측치는 연도가 지날수록 많아지는 경향을 보이는데, 이는 상장 폐지 기업보다 신규 상장 기업이 더 많은 데 따른 것으로 추측된다. 본 연구에 쓰인 분석 표본의 연도별 분포는 아래 표 1과 같다. 표본 내 데이터(In-sample)는 2003~2014년, 표본 외 데이터(Out-sample)는 2015~2019년으로 설정하였다.

표 1. 연도별 표본 분포

연도	기업수	비율(%)	누적비율(%)
2003	318	4.86	4.86
2004	317	4.84	9.70
2005	335	5.12	14.82
2006	343	5.24	20.06
2007	375	5.73	25.79
2010	407	6.22	32.01
2011	437	6.68	38.69
2012	448	6.84	45.53
2013	455	6.95	52.48
2014	465	7.10	59.59
2015	492	7.52	67.10
2016	507	7.75	74.85
2017	528	8.07	82.92
2018	554	8.46	91.38
2019	564	8.62	
계	6,545	100.00	100.00

Ⅲ-2. 설명변수

본 연구는 Amini et al.(2021)를 참고하여 재무·거시경제 변수를 산출하였다. 기업의 재무 데이터는 FN 가이드의 데이터가이드에서, 거시경제 변수는 한국은행 경제통계시스템 홈페이지에서 입수하였다. 법인세율, 알트만 Z-점수 등은 회계법인 홈페이지와 관련 논문을 참고해 작성하였다. 구체적으로 종속(재무) 변수는 시장가치총부채비율(TDM), 총자산총부채비율(TDA), 시장가치장기부채비율(LDM), 총자산장기부채비율(LDA)이며, 설명 재무 변수는 연도말 시총(MVE), 자산의 시장가치(MVA), 총자산대비 영업이익(Profit), 총자산 변화율(Assets), 기업 장기존속 여부(Mature), 시장가-장부가 비율 (Mktbk), 총자산 증감률(ChgAsset), 자본지출(CAPEX), 유형자산(Tang), 연구개발비(RD), 특수산업(Unique), 판관비(SGA), 현금성자산(Cash), 법인세율(Taxrate), 감가상각비(Depr),

주가변동성(StockVar), 알트만 Z-점수(Zscore), 신용등급(Rating), 개별 주가수익률(Stock), 시장 주가수익률(Crspret), 부채비율 중앙값(indstlev), 총자산 증감률 중앙값(industgr), 10 년과 1 년물 국채스프레드 수익률 차이 (termsprd), 예상물가상승률(inflation), 당기순이익 증감율(Macroporf), 실질 GDP 성장률(Macrogr), 순배당율(Netpay), 자산 더미 변수(Size), 기업 더미(Growth), 첨단기술 더미(Hightech)를 활용하였다. 본 연구에 활용한 재무·거시경제 변수와 산출 방법은 아래 표 2 와 같다.

표 2. 재무·거시경제 변수

변수명	산출 방법
MVE	기업 주식 종가와 보통주식수의 곱
MVA	유동부채 + 장기부채 + 우선주 청산가치 – 이연법인세 + MVE
TDM	(유동부채 + 장기부채) /시가총액
TDA	(유동부채 + 장기부채) /총자산
LDM	장기부채 / 시가총액
LDA	장기부채 / 총자산
Profit	(감가상각전) 영업이익 / 총자산
Assets	총자산 로그값
Mature	기업 데이터가 최근 5 년 이상 존재 시 1, 5 년 이하 시 0
Mktbk	시가총액 / 총자산
ChgAsset	총자산 (연간) 증감률
Capex	자본지출 / 총자산
Tang	유형자산 / 총자산
RD	연구개발비 / 총매출
Unique	우주선, 유도미사일, 비행기, 컴퓨터, 반도체, 화학 관련 업종 시 1, 그 외 0
SGA	판관비 / 총매출
Cash	현금 및 단기투자자산
Taxrate	법인세 최고세율
Depr	감가상각비 / 총매출
StockVar	일일 주가수익률의 연간변동성
Zscore	알트만의 Z-점수 [2]

[2] 신흥국의 경우 파산비율을 뜻하는 Z-점수 계산 방법은 아래와 같다. Meeampol et al.(2014)을 참고함.

Rating	회사별 신용등급 BB 이상 1, 신용등급 BB 미만 0
Stock	연간 주식누적수익률
CrspRet	주식시장의 연간누적수익률
Industlev	TDM 중앙값
industgr	ChgAsset 중앙값
Termsprd	10 년물 국채와 1 년물 국채의 수익률 차이
inflation	예상물가상승률
Macroporf	비제조 기업의 연간 당기순이익 로그값
Macrogr	실질 GDP 성장률
Netpay	(현금배당금 + 보통주 및 우선주 순증액) / 총자산
Size	총자산의 상위 30% 기업이면 2, 중간 40%면 1, 하위 30%면 0
Growth	Mktbk 변수 크기가 상위 30% 기업이면 2, 중간 40%면 1, 하위 30%면 0
High-Tech	IT 제품 및 서비스 기업이면 1, 그 외면 0

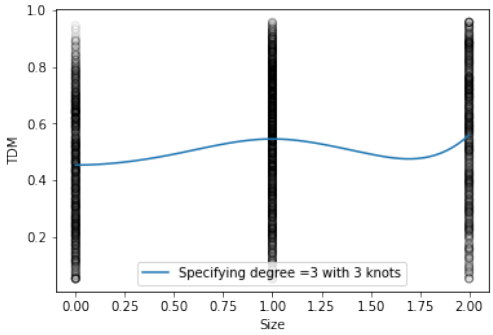
이어 본 연구에서 제시한 종속 변수와 각 재무 및 거시경제 변수의 관계는 비선형성 및 내생성을 가질 수 있다(Graham and Leary, 2011). 이와 같은 관계를 확인하기 위하여 본 연구는 스플라인 보간법(Cubic Spline Interpolation)을 분석할 것이다. 또한 선행 연구(Amini et al., 2021; Graham et al, 2011)가 꼽은 기업 자산 더미(Size), 수익성(Profit), 유형자산(Tangibility), 시장가 대비 장부가 비율(Market-to-Book ratio), 알트만 Z-점수(Z-score), 현금성 자산(Cash) 등 재무 변수와 대표 부채비율 변수인 시가총액 대비 부채비율(TDM)의 관계는 아래 그림 1에 표현되어 있다. 본 연구는 선행연구를 참고하여 표본 내 데이터를 활용하여 추정치를 구하였다. 이 그림에 따르면 각 목표 부채비율로의 결정 요인들과 주요 종속변수인 시장가치총부채비율(TDM) 간의 관계는 비선형(non-linear) 형태로 나타나고 있다는 점을 알 수 있다. 즉, 본 연구에서 머신러닝 모형을 활용한다면 결정 요인과 부채비율 변수 간에 사전 포착되기 어려운 비선형성과 내생성 문제를 해결할 수 있을 것으로 보인다. 이러한 문

$Z = 3.25 + 6.56X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4$

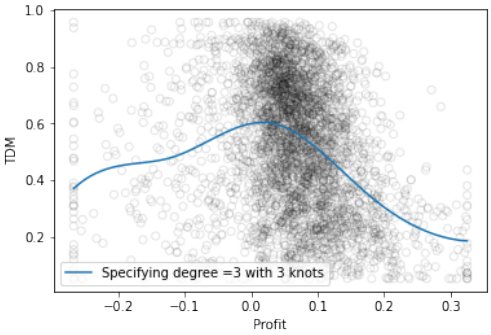
X1 = (유동자산 – 유동부채) / 총자산
X2 = 유동이익/총자산
X3 = EBIT/총자산
X4 = 총자본/총부채

제는 기업의 자금 조달 결정에 고유한 것으로 알려져 있다(Childs et al., 2005).

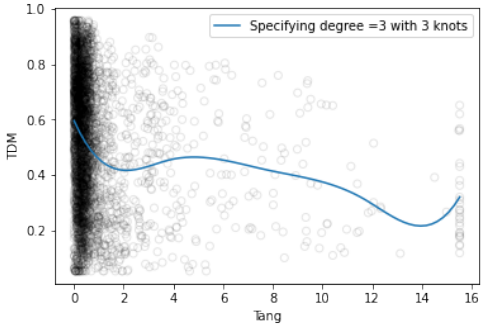
그림 1. 기업 부채비율과 주요 변수 간 관계



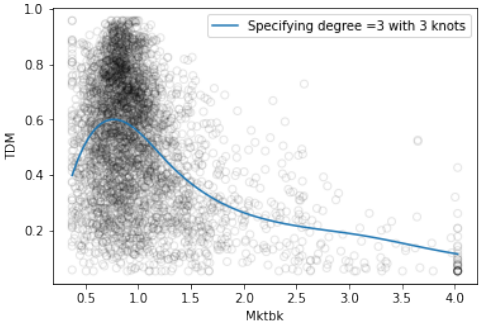
자산 규모



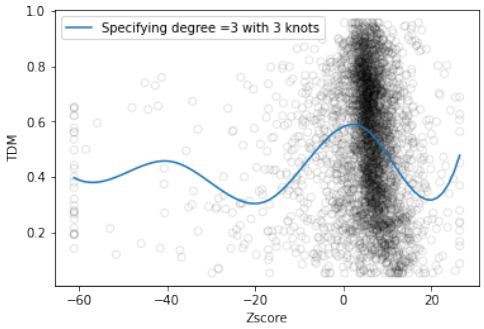
수익성



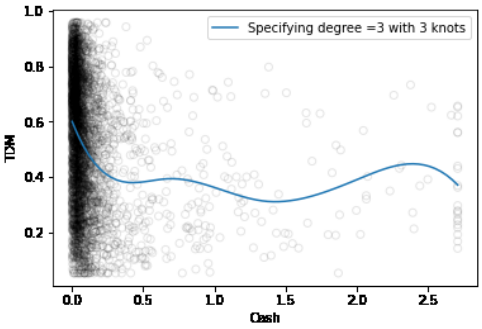
유형자산



시장가 대비 장부가 비율



알트만 Z-점수



현금성 자산

Ⅲ-3. 분석모형

본 연구는 최적의 부채비율을 예측하기 위해 사용한 실증 모델을 제시한다. 먼저 머신러닝 모델을 사용하여 목표 부채비율의 예측 성능을 개선하고, 목표 부채비율로의 조정 속도를 추정한다. 최적 부채비율을 예측하는 기본 회귀식은 함수 $g(X_{i,t}) = E(y_{i,t+1} \mid X_{i,t})$를 따른다. 여기에서 $y_{i,t+1}$는 다음과 같이 정의된다.

$$y_{i,t+1} = g(x_{i,t}) + \varepsilon_{i,t+1} \ldots (1)$$

$y_{i,t+1}$는 t+1년에 i번째 기업의 목표 부채비율을 말하며, $\varepsilon_{i,t+1}$은 임의 오차 성분을 말한다. 따라서 회귀 함수 $E(y_{i,t+1} \mid X_{i,t})$는 공변량 벡터를 조건으로 하는 $y_{i,t+1}$의 조건부 기대값이다. 본 연구의 목표는 LM, LASSO, RF, GBM 등 다양한 방법을 사용하여 함수 $g(X_{i,t})$를 추정하는 것이라고 할 수 있다.

Ⅲ-3-1. 선형 모형

Ⅲ-3-1-1. 다중회귀 모형

본 연구의 기본 모형은 다음과 같은 회귀식을 취하고 있다.

$$g(X_{i,t}; \beta) = X'_{i,t}\beta \ldots (2)$$

위 식 (2)에서 최소자승법(OLS)을 사용해 추정한 매개변수 β는 다음과 같이 정의된다.

$$\hat{\beta}^{ols} = \operatorname{argmin} \| y - x\beta \|_2^2 \ldots (3)$$

다시 위 식 (3)에서 $\| a - b \|_2$는 a와 b 벡터의 거리를 뜻한다.

Ⅲ-3-1-2. 라쏘 모형

본 연구에서 다중 회귀 모형을 사용하는 데는 과적합(Overfitting)과 잠재적인 다중 공선성(Potential multi-collinearities)이라는 한계가 존재할 수 있다. 이에 본 연구는 라쏘(LASSO, Least Absolut Shrinkage and Selection Operator)을 활용하고자 한다 (Tibshirani, 1996). 라쏘는 덜 중요한 공변량과 관련된 매개변수를 0으로 축소하는 효과적인 방법으로 꼽힌다. 결과적으로 라쏘는 목표 부채비율을 예측하는데 있어 중요한 변수를 활용함으로써 최적의 자본구조를 구현하는데 필요한 변수를 고를 수 있다. 다음과 같이 주어진 λ 값에 대한 라쏘 모델의 매개 변수 추정값은 다음과 같다.

$$\hat{\beta}_{\lambda}^{\,LASSO} = \operatorname{argmin}\|y\text{-}x\beta\|_2^2 + \lambda\|\beta\|_1 \text{ for some } \lambda > 0 \ldots (4)$$

Ⅲ-3-2. 머신러닝(Machine Learning) 모형

머신러닝 알고리즘은 종속 및 독립 변수 간의 비선형 관계를 모델링하고 예측하는 데 유용하게 쓰일 수 있다. 특히 변수 간의 숨겨진 상호 작용. 그리고 비선형 관계 및 상호 작용 효과를 분석하는데 탁월한 성능을 보이고 있다. 대조적으로, 여기에 설명된 ML 방법은 완전히 비모수적이며 충분히 사전 개입 없이 이러한 복잡한 구조를 유연하게 포착할 수 있다. 본 연구에서 활용한 머신러닝 모델은 랜덤 포레스트 (Breiman, 2001)와 GBM(Friedman, 2001)이다.

Ⅲ-3-2-1. 랜덤포레스트(RF) 모형

랜덤포레스트는 앙상블 계열 모델로 앙상블 머신러닝 모형이다. 여러 개의 의사결정 트리를 형성한 이후에 새로운 데이터를 각 트리에 분류한 최종 분류 결과를 선택한다. 구체적으로 공간 $\mathscr{X}$를 **J**의 고유하면서 겹치지 않는 영역: R_1, R_2, ..., R_J 내의 모든 값에 대한 y의 예측 값의 전체 평균치를 말한다. 이를 수식화시키면 다음 식 (5)와 같다.

$$\hat{g}^{\,rf}(x) = \sum_{j=1}^{J} \bar{y}_j \, I\{x \in R_j\} \ldots (5)$$

여기서 $I_{\{x \in Rj\}}$는 x가 R_j에 포함되면 1, 그렇지 않으면 0을 뜻하는 지시 함수이다. 본 모형에서 한 개의 의사결정 트리는 높은 분산으로 인하여 샘플이 바뀔수록 불안정해질 수 있다. 따라서 본 연구는 Efron et al.(1994)를 참고하여, '배깅'이라고 불리는 부트스트랩 방법을 활용하였다. 부트스트랩은 데이터의 샘플링 방식을 말하는데, 각 모형이 서로 다른 훈련 데이터를 이용하고, 그 데이터 세트를 추출할 때 복원 추출하며, 원 데이터 수만큼 데이터 셋을 뽑는다는 특징을 가지고 있다. 트리 상에서의 결과값을 합친 결과 결정 트리의 앙상블이 생성되는데, 여기서 x의 배깅 추정치는 모든 트리에 대한 평균 추정치를 말한다. 이에 대한 수식은 다음 식 (6)과 같다.

$$\hat{g}^{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{g}^{b}(x) \quad \ldots (6)$$

여기서 $\hat{g}^{bag}(x)$는 식 (6)에 따라 정의된 추정량이다. b번째 부트스트랩 샘플에서 이 앙상블에 대한 평균화의 효과는 최종적인 추정치의 변동을 줄이는 것이다(Breiman, 1996).

Ⅲ-3-2-1. GBM(Gradient Boost Regression) 모형

GBM 모형은 회귀 및 분류 분석을 수행하는 예측 모형으로 ML 모형 중 안정적인 예측 성능을 가진 것으로 알려져 있다. 다양한 분류기를 결합하여 더 높은 정확도를 자랑하는 부스팅 계열의 앙상블 알고리즘이다. 특히 GBM 모형은 반복적인 잔차 적합(Residual fitting) 과정을 통하여, 개별 추정치의 가중치 합인 $\hat{g}_{gbm}(X)$를 구할 수 있다. 여기서 가중치는 모델이 학습하는 속도를 결정하는 매개변수 λ에 의해 제어되는 동시에, 교차 검증을 통해 발견된다(Amini et al., 2021).

Ⅲ-3-3. 그리드 서치를 통한 하이퍼파라미터 튜닝

각 선형 및 머신러닝 모형 분석에 앞서 본 연구는 그리드 서치(격자 탐색)를 통해 하이퍼 파라미터 조합을 정하였다. 그리드 서치란 순차적으로 하이퍼 파라미터 값을 입력한 뒤에 가장 높은 성능을 보이는 최적의 하이퍼 파라미터들을 찾는 탐색 방법이다. 이와 같은 방법을 통해 튜닝된 매개변수의 성능은 훈련된 모델이 검증 세트의 목표 부채비율을 얼마나 잘 예측하는지 여부로 평가할 수 있다.

Ⅵ. 분석 결과

Ⅵ-1. 평균 제곱 오차 및 결정계수(R^2_{OS})분석

본 연구는 그리드 탐색을 마친 각 선형 및 머신러닝 모델을 활용, 표본 외 데이터(2015~2019년)에 대하여 회귀 문제의 전형적인 성능 지표인 평균 제곱 오차(Mean square error)과 결정계수(R^2_{OS})[3]를 연도별로 예측하였다. 본 연구의 대표 종속 변수인 시장가치총부채비율(TDM)에 대하여 모델별 평균 제곱 오차[4]와 결정계수(R^2_{OS})를 예측한 결과는 아래 표 3과 표 4에 정리했다.

표 3. 기업의 부채비율을 예측한 R^2_{OS}

	2015년	2016년	2017년	2018년	2019년	2015~2019년
다중회귀	-0.105230	0.031941	0.401143	0.126228	-1.609892	-0.265298
라쏘	0.006570	0.103022	0.402515	0.191566	-1.210615	-0.127837
랜덤포레스트	0.606779	0.565904	0.523065	0.532053	0.527210	0.552908
GBM	0.591878	0.545335	0.404715	0.537237	0.543981	0.528520

[3] 여기서 R^2_{OS}은 표본 외 데이터(Out-of-sample)에 대한 결정 계수라는 의미이다.

[4] 평균 제곱 오차는 회귀 문제의 전형적인 성능 지표로 오차가 커질수록 이 값이 커짐에 따라 예측에 얼마나 많은 오류가 있는지 가늠하게 해준다.

표 4. 기업의 부채비율에 대한 평균 제곱 오차

	2015년	2016년	2017년	2018년	2019년	2015~2019년
다중회귀	0.059471	0.049278	0.031259	0.047088	0.149441	0.068474
라쏘	0.053456	0.045660	0.031187	0.043567	0.126579	0.061035
랜덤 포레스트	0.021159	0.022097	0.024895	0.025218	0.027072	0.024195
GBM	0.021961	0.023144	0.031072	0.024939	0.026111	0.025515

먼저 표 3은 2015년부터 2019년까지 표본 외 데이터 결정계수(R^2_{OS})를 나타내고 있다. 선형 모형과 라쏘는 각각 -0.16~0.40, -1.21~0.40의 예측력을 보이는 한편, 랜덤 포레스트와 GBM은 각각 0.52~0.61, 0.40~0.59의 우월한 성능을 나타내고 있다. 이는 선형 모형과 비교하여 랜덤 포레스트와 GBM의 예측력이 더욱 높다는 점을 보여주고 있다. 앞서 다중회귀 모형에서 낮은 R^2_{OS} 값은 매개 변수 불안정으로 인한 모형 실패인 것으로 보인다(Amini et al, 2021). 이어 표 4는 모형 별 평균 제곱 오차 값을 보여주고 있다. 선형 모형인 다중회귀 모형과 라쏘의 평균 제곱 오차는 각각 0.03~0.15, 0.13~0.05를 기록한 반면, 랜덤 포레스트와 GBM은 0.02~0.03, 0.02~0.03으로 평균 제곱 오차가 상당히 낮은 수준임을 나타내고 있다. 요약하면, 랜덤 포레스트와 GBM은 다중회귀, 라쏘 등 선형 모델에 비해 부채 비율에 대한 예측력이 높으며 이에 대한 예측 정확도 역시 높다는 점을 알 수 있다. 이는 기존 선형 모형의 높은 평균 제곱 오차에서 볼 수 있듯이 목표 부채비율에 대한 예측력이 상당히 떨어진데 따른 것으로 추정할 수 있다.

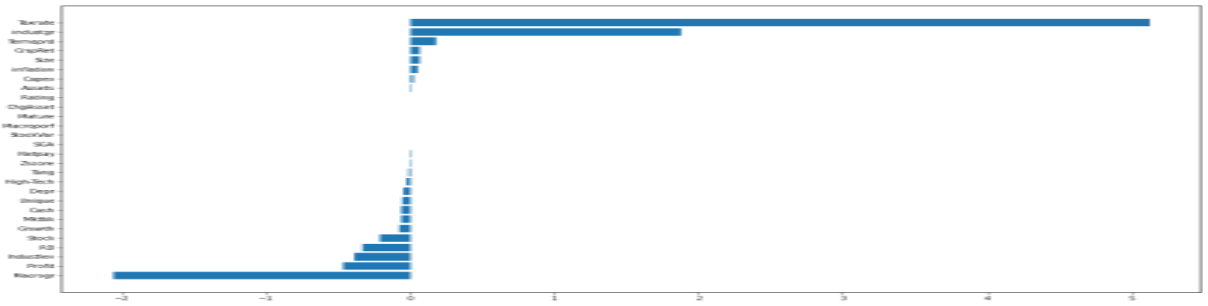
VI-2. 변수 중요도 추정

이어 본 연구는 모형별로 분석한 변수 중요도(Variable Importance)를 확인하고자 한다(Amini et al, 2021; Strobl et al, 2008). 변수 중요도란 예측 성능에 중요한 역할을 한 변수를 추정하는 것을 말한다. 구체적으로 본 연구는 종속 변수인 시장가치총부채비율(TDM)에 대한 선형 및 머신러닝 모형별 변수 중요도를 분석하였다. 그 결과는

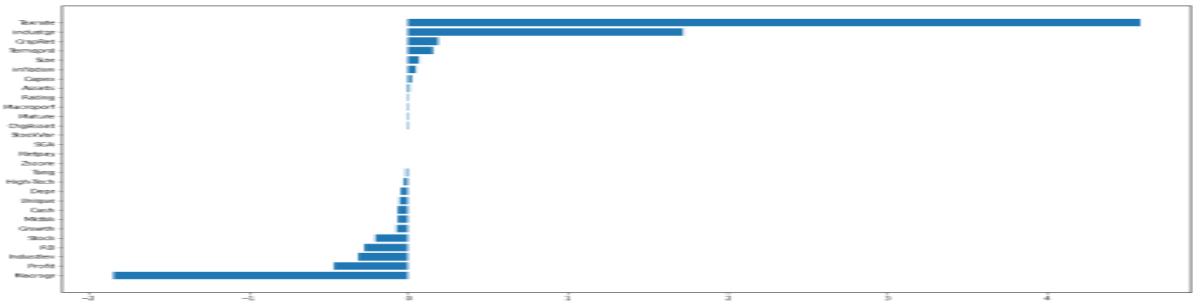
아래 그림 2에 표현하였다.

이 그림에 따르면 선형 모형인 다중회귀와 라쏘, 머신러닝 모형인 GBM에서 부채비율 예측에 대하여 설명력이 높은 변수는 법인세율, 총자산증감률 중앙값, 국채스프레드, 시장의 주가수익률 등으로 대부분 경제 변수이거나 재무 변수의 중간값이며, 상당수 재무 변수는 음수를 기록하고 있다는 것을 알 수 있다. 다시 말해, 선형 모형은 설명력이 높은 기업의 개별 재무 변수를 갖고 있지 않다는 점을 알 수 있다. 반면 랜덤포레스트의 경우 , 순배당율(Netpay), 알트만 Z-점수, 영업이익(Profit), 총자산 더미(Growth), 현금성 자산(Cash) 순으로 기업의 재무 변수의 예측력이 높은 것으로 나타난다. 다시 말해, 랜덤 포레스트 모형의 기업 재무 변수는 다른 모형에 비하여 매개변수가 안정적이고 예측력이 높다는 사실을 알 수 있다.

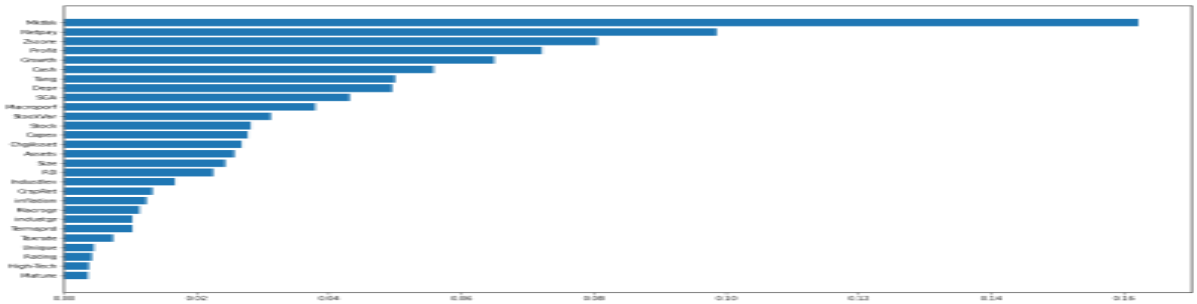
그림 2. 시장가치총부채비율에 대한 선형 및 머신러닝 모형의 변수 중요도



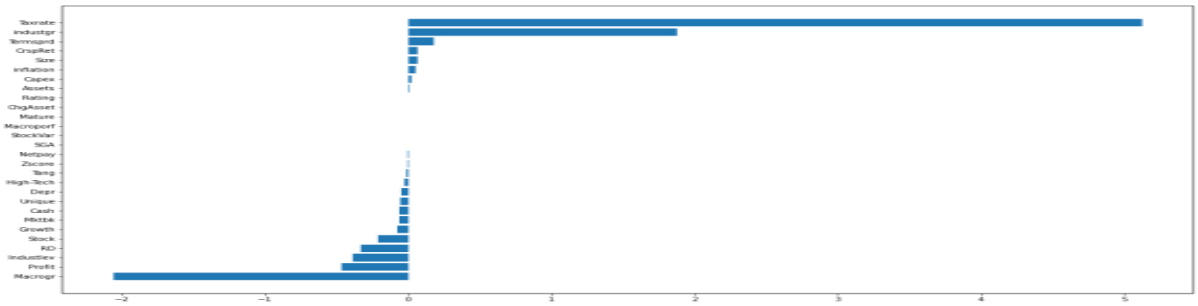
LM



LASSO



RF



GBM

VI-3. 목표 부채비율로의 조정 속도 추정

머신러닝(ML)은 비선형적이거나 내생성이 높은 데이터에 대하여 성능이 뛰어난 것으로 알려져 있다(Altman et al, 2017; Strobl et al., 2008). 기존의 상당수 연구는 목표 부채비율을 '추정'하는데 그치는 반면, 머신러닝은 훈련 세트에서 '적합치(fitted value)'를 추정한 뒤 이를 기반으로 예상 값을 예측하기 때문이다. 본 연구는 기존 선형 모형과 ML 모형을 활용, 표본상 기업들이 실제 부채비율과 목표 부채비율의 격차를 얼마나 조정(감소)시킬 수 있는지 비교 분석하고자 한다. 실제 부채비율과 목표 부채비율의 차이는 $GAP_{i,t} = E(y_{i,t+1})-y_{1,t}$이며, 이에 따른 목표 부채비율 조정 속도 추정 모형은 다음 식 (7)과 같다.

$$\triangle y_{i,t+1}=\lambda GAP_{i,t} + \varepsilon_{i,t+1} \ \ldots (7)$$

식 (7)에서 λ는 기업이 t년 동안 목표 부채비율을 향해 조정하는 속도(비용)를 말한다. 만약 기업 경영자가 목표 부채비율로 도달하기 위해 노력을 기울인다면 λ 값은 0을 넘겨야 한다(Amini et al, 2021). 다만, 시장 마찰이 있다면 부채 조정은 즉각적으로 이뤄지지 않을 수 있으며, 이에 따라 λ는 1보다 낮은 숫자를 나타낼 수 있다.[5] 이어 본 연구는 목표 부채비율에 대한 기업의 조정 속도를 분석하기 위하여 식 (7)에 따라 λ로 측정한 기업의 부채비율 조정 속도(SOA)와 반감기(half-life)를 추정한다. 대표 종속 변수는 시장가치총부채비율(TDM), 총자산총부채비율(TDA)을 활용하였으며, 패널 A는 시장가치총부채비율, 패널 B는 총자산총부채비율에 대하여 분석한 것이다. (1)~(4)열의 다중회귀, 라쏘, 랜덤 포레스트, GBM 모형 분석이며, (5)~(8)열은 기업 고정 효과(firm-fixed effect)를 추가한 것이다. 그 결과는 아래 표 5와 같다.

[5] 이는 앞서 언급한 상충이론과도 비슷한 맥락의 접근이라고 할 수 있다. 윤보현 외(2016)는 부채비율 조정속도가 1로 추정된다면 현실 기업의 자본구조가 절충이론에 완전히 합치한다는 것을 의미하며, 0으로 추정된다면 상충이론이 현실 기업에 대해 전혀 설명력을 갖지 못한다는 것을 의미한다는 점을 제시하고 있다.

표 5. 기업의 목표 부채비율로의 조정 속도 추정

	기업 고정 효과 제거				기업 고정 효과 포함			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
패널 A: 시장가치총부채비율	다중회귀	라소	랜덤 포레스트	GBM	다중회귀	라소	랜덤 포레스트	GBM
부채비율 조정속도	0.07*** (0.006)	0.08*** (0.006)	0.11** (0.011)	0.11** (0.011)	0.14** (0.013)	0.16** (0.013)	0.45** (0.024)	0.32** (0.022)
관측치	2,650	2,650	2,650	2,650	2,650	2,650	2,650	2,650
조정 결정계수	0.056	0.057	0.041	0.036	0.056	0.064	0.145	0.091
반감기	3.78	3.69	3.16	3.23	2.81	2.61	1.16	1.64
패널 B: 총자산대비총부채비율								
부채비율 조정속도	0.05*** (0.008)	0.06*** (0.008)	0.08*** (0.01)	0.08** (0.011)	0.28** (0.02)	0.32** (0.021)	0.40** (0.024)	0.30** (0.022)
관측치	2,650	2,650	2,650	2,650	2,650	2,650	2,650	2,650
조정 결정계수	0.017	0.018	0.022	0.019	0.053	0.067	0.087	0.041
반감기	4.27	4.18	3.62	3.71	1.82	1.63	1.31	1.74

Symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

표 3의 (1)~(4)열 결과를 요약하면 다음과 같다. 첫째, 머신러닝 모형의 부채비율 조정속도는 선형 모형에 비해 더욱 빠른 것으로 나타나고 있다. 구체적으로 패널 A의 시장가치총부채비율을 살펴보면 다중회귀와 라쏘는 통계적으로 유의한 수준에서 7%, 8%을 보인 반면, 랜덤포레스트와 GBM은 각 11%로 빠른 조정 속도를 나타내고 있다. 패널 B의 총자산대비총부채비율에 대하여 다중회귀와 라쏘는 각 5%, 6%를 보인 반면, RF와 GBM은 각 8%의 조정 속도를 나타내고 있다. 요약하면 패널 A와 패널 B의 부채비율에 대하여 머신러닝 모형은 선형 모형에 비하여 50%가량 높은 조정 속도로 목표 부채비율에 수렴하고 있다. 또한 패널 A에서 다중회귀와 라쏘로 추정한 목표 부채비율로의 반감기는 3.78, 3.69가 나온 반면, 랜덤포레스트와 GBM은 3.16, 3.23을 나타냈으며, 패널 B에서 다중회귀와 라쏘로 추정한 목표 부채비율로의 반감

기(half-life)는 4.27, 4.18가 나온 반면, 랜덤포레스트와 GBM은 3.62, 3.71을 나타내는 등 머신러닝 모형의 속도가 선형 모형에 비하여 더욱 빠르다는 점을 알 수 있다.

이어 (5)~(8)열의 기업 고정 효과를 추가한 각 모형에서도 머신러닝 모형은 훨씬 더 개선된 성능을 보이고 있다. 패널 A에서 랜덤 포레스트와 GBM의 목표 부채비율 조정 속도는 각 0.45, 0.32로 기업 고정 효과를 뺀 모형에 비하여 3~4배가량 증가하였다. 다중회귀와 라쏘의 목표 부채비율 조정 속도가 0.14, 0.16으로 소폭 증가한 것과 비교하면 대조적이다. 패널 B에서도 랜덤 포레스트와 GBM의 목표 부채비율 조정 속도는 각 0.40, 0.30으로 기업 고정 효과를 뺀 모형에 비하여 4배가량 증가하였다. 반면 다중회귀와 라쏘는 0.28과 0.32를 기록하며 시장가치총부채비율에 비하여 높은 증가 폭을 기록하였다. 그러나 반감기는 랜덤 포레스트와 GBM 모형에 비하여 여전히 낮은 수준을 보였다.

이와 같은 결과는 목표 부채비율 조정 속도를 예측한 최근 연구 결과(Amini et al., 2021)와 대체적으로 일치한 양상을 나타내고 있다. 구체적으로 최근 선행 연구인 Amini et al.(2021)와 비교하면 시장가치총부채비율(패널 A)로의 조정 속도를 비교했을 때 본 연구의 랜덤 포레스트는 0.11을 기록해 선행 연구의 0.215보다 낮았지만, 기업 고정 효과를 포함시키면 0.45를 기록, 선행 연구 수준(0.473)과 비슷한 수준이었다. 총자산대비총부채비율(패널 B)로의 조정 속도를 보면 본 연구의 랜덤 포레스트는 기업 고정 효과가 없으면 0.08을 기록해 0.150을 기록한 선행 연구에 비해 낮았지만 기업 고정 효과를 포함시키면 0.40으로 선행 연구의 0.345보다 더 높은 것으로 나타났다.

이어 본 연구의 목표 부채비율 조정 속도는 전통적인 모형을 쓴 기존 연구와 비슷하거나 높은 수준을 기록했다. 김영래 외(2007)에서 일반고정효과모형을 썼을 때 시장가치총부채비율에 대한 부채비율 조정 속도는 0.35를 기록했으며, 윤보현 외(2016)에서 최소자승법으로는 0.228, Fama et al.(1973) 모형은 0.229을 기록했다. 따라서 세부 변수는 연구별로 다르다 하더라도 본 연구의 머신러닝 기법으로 추정한 목표 부채비율로의 조정 속도 성능이 상대적으로 더욱 높다는 점을 확인할 수 있다.

Ⅶ. 결론

본 연구의 주요 분석 결과는 다음과 같다. 첫째, 각 머신러닝 모델을 활용한 결과 본 연구는 랜덤포레스트와 GBM을 활용한 기업의 목표 부채비율에 대한 조정 속도 성능이 다중회귀, 라쏘에 비하여 50%가량 높은 것을 확인하였다. 둘째, 랜덤포레스트와 GBM 모델을 활용한 결과 변수 중요도(Variable Importance)를 분석한 결과, 시장가-장부가 비율 (Market-to-Book ratio), NeyPay(순발행액), Z-Score, Profit(매출액) 순인 것으로 나타났다. 셋째, 각 모델별로 목표 부채비율로의 부채비율 조정 속도를 분석한 결과, 앞서 분석과 마찬가지로 랜덤포레스트, GBM, 다중회귀, 라쏘 순으로 부채비율 조정 속도가 빠른 것으로 나타난다.

본 연구의 대표적인 공헌점은 다음과 같다. 첫째, 국내 코스피 시장에 상장된 이력이 있는 기업 전체를 대상으로 목표 부채비율을 예측하고 이로의 조정 속도를 예측하는 머신러닝 분석을 실시하였다는 것이다. 특히 선형 모형과 머신러닝 모형을 활용하여 모형 별 성능을 비교 분석함으로써 목표 부채비율과 관련된 국내 연구의 외연을 넓혔다. 둘째, 다양한 재무·거시경제 환경에서의 머신러닝 분석 기법의 이점을 증명하였다는 것이다. 특히 머신러닝 모형이 과적합성, 내생성과 관련된 여러 재무 연구에 폭넓게 쓰일 것으로 기대된다. 한편 본 연구는 국내 상장기업을 다양한 산업군으로 분석을 실행하지 못하고, 제한된 데이터로 인하여 더욱 다양한 함의를 찾지 못 하였다는 점이 아쉬움으로 남는다.

Ⅷ. 참고문헌

김영래·김필규·최종범, “자본구조 결정요인과 부채비율 조정속도에 관한 연구,”

2007 5개 학회 공동학술연구 발표회.

김진수, “자본조달시장접근성과 자본구조조정속도,”『재무연구』제23권 제2호, 2010, 89-120.

박연희·최효순·김한수, “재량적발생액이 회사채수익률 스프레드에 미치는 영향,”『회계저널』제20권 제4호, 2011, 35-56

손승태·이윤구, “코스닥기업의 자본구조 결정요인, 동태적 자본구조 모형을 중심으로,”『재무관리연구』제24권 1호 2007.3. 109~147

손판도·손승태, “자본구조의 평균회귀현상과 장기균형,”『재무관리연구』제25권 제3호, 2008, 33-78

신민식·김수은(2008), “기업의 소유구조와 기업가치간의 관계,”『금융지식연구』wp9권 제2호, 2011, 129-157

여희정, “해운기업의 목표레버리지와 레버리지 결정요인,”『무역학회지』 제43권 제2호, 2018, 181-204

윤보현·이정환·하준·손삼호, “기업재무 데이터특성을 고려한 자본구조 조정속도 추정에 관한 연구,”『지역산업연구』제39권 제2호, 2016, 55-84

윤봉한·오재영, “기업지배구조와 기업성과 및 기업가치: 한국상장기업에 대한 실증연구”『한국증권학회지』, 2005 제34권 1호, 227-263

이원흠·이한득·박상사, “대기업집단의 부채비율 조정속도에 관한 연구”,『한국증권학회지』, 2001, 87-114

최윤이, “금융위기 전후 부채비율이 기업가치에 미치는 영향,”『산업연구』제39권 제1호, 2015, 35-59

Altman, N. and Krzywinski, M., “Ensemble methods: bagging and random forests,” Nat. Methods 14, September 2017, 933–934.

Cristina Aybar-Aria, Alejandro Casino-Martínez and José López-Gracia, “On the

adjustment speed of SMEs to their optimal capital structure," *Small Business Economics* 39, November 2012, 977–996.

Efron, B. and Tibshirani, R.J. "An Introduction to the Bootstrap" Chapman and Hall/CRC, UK. 1994.

Erwan Morellec, "Can Managerial Discretion Explain Observed Leverage Ratios?" *The Review of Financial Studies* 17, January 2004, 257-294.

Eugene F. Fama and Kenneth R. French, "Testing Trade-Off and Pecking Order Predictions about Dividends and Debt," *The Review of Financial Studies* 15, Spring 2002, 1-33.

Eugene F. Fama and James D. MacBeth, "Risk, return, and equilibrium: empirical tests," *The Journal of Political Economy* 81, May-June 1973, 607-636.

Murray Z. Frank and Vidhan K. Goyal Frank, "The effect of market conditions on capital structure adjustment," *Finance Research Letters* 1, March 2004, 47-55.

Giovanni S.F. Bruno, "Approximating the bias of the LSDV estimator for dynamic unbalanced panel data models," *Economics letters* 87, June 2005, 361-366.

Michael C. Jensen, "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers", *The American Economic Review* 76, May 1986, 323-329.

Michael C. Jensen and William H. Meckling, "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of Financial Economics* 3, October 1976, 305-360.

Jerome H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine", *The Annals of Statistics* 29, October 2001, 1189-1232.

John R. Graham and Campbell R. Harvey, "The theory and practice of corporate finance: evidence from the field," *Journal of Financial Economics* 60, May 2001, 187~243.

JR Graham and CR Harvey, "The theory and practice of corporate finance: Evidence from the field," *Journal of Financial Economics* 60, May 2001,187-243.

Julio Pindado and Alberto de Miguel, "Determinants of Capital Structure: New Evidence from Spanish Panel Data," *Journal of Corporate Finance* 7, December 2001, 77-99.

Lakshmi Shyam-Sunder and Stewart C. Myers, "Testing static tradeoff against pecking order models of capital structure," *Journal of Financial Economics* 51, February 1999, 219-244.

Leo Breiman, “Bagging predictors,” *Machine Learning* 24, August 1996, 123–140.

Leo Breiman, “Random Forests,” *Machine Learning* 45, October 2001, 5–32.

Malcolm Baker and Jeffrey Wurgler, “Market Timing and Capital Structure,” *The Journal of Finance* 57, December 2002, 1-32.

Mark Flannery and Kasturi P. Rangan, “Partial adjustment toward target capital structures,” *Journal of Financial Economics* 79, March 2006, 469-506.

Murray Z. Frank and Vidhan K. Goyal, “Capital Structure Decisions: Which Factors Are Reliably Important?” *Financial Management* 38, Spring 2009, 1-37

Paul D.Childs, David C.Mauer and Steven H.Ott., “Interactions of corporate financing and investment decisions: The effects of agency conflicts,” *Journal of Financial Economics* 76, June 2005, 667-690.

RenéM.Stulz∗, “Managerial Discretion and Optimal Financing Policies”, *Journal of financial Economics* 26, 1990, 3-27.

Robert Tibshirani, “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society* 58, July 1990, 267-288.

Sasivimol Meeampol, Polwat Lerskullawat, Ausa Wongsorntham, Phanthipa Srinammuang, Vimol Rodpetch and Rungsimaporn Noonoi, “Applying Emerging Market Z-Score Model To predict Bankruptcy: A case study of Listed companies In The Stock Exchange Of Thailand (Set),” Management, Knowledge and learning International conference 2014.

Shahram Amini, Ryan Elmore, Ozde Oztekim and Jack Strauss, “Can machines learn capital structure dynamics?” *Journal of Corporate Finance* 70, October 2021, 1–22.

Stewart C. Myers, “The Capital Structure Puzzle,” *The Journal of Finance* 39, July 1984, 574-592.

Strobl, C., Boulesteix, A., Kneib, T., Augustin, T. and Zeileis, A., “Conditional variable importance for random forests,” *BMC Bioinform* 9, July 2008, 307–317.

Can Machine Learning Predict Capital Structure Dynamics?

- Evidence from Korea

JeongHwan Lee*, JinHyung Cho**, BongJun Kim***,WonEung Lee****

A B S T R A C T

Yes, it can. Employing a variety of Machine Learning (ML) algorithms, we predict optimal capital structure of listed firms in Korea, comparing the performance of linear and machine learning models - namely, Multi-regression, LASSO, Random Forest (RF) and Gradient Boosting Regression (GBM). For analysis, we set the training and test set as 2003-2014 and 2015-2019 respectively. We find that the predicting performance on firm leverage, as measured in out-R^2_{OS} and MSE (Mean Square Error) for RF and GBM is much effective than that of LM and LASSO. In particular, the variables with high predictive power are the Market-to-Book ratio, NetPay, Zscore, Profit, and so on. Finally, after estimateing the speed of adjustment (SOA) to the optimal capital structure, using the model of Amini et al. (2021), we confirm that RF and GBM are around 50% more predictive than LM and LASSO.

Keyword: Machine Learning, Korea, Leverage, Leverage Adjustment Speed, Optimal Capital Structure

*First Author, Associate Professor, College of Economics and Finance, Hanyang University, jeonglee@hanyang.ac.kr

**Corresponding Author, Phd Candidate, College of Economics and Finance, Hanyang University, enish27@hanyang.ac.kr

***Co-Author, Phd Candidate, College of Economics and Finance, Hanyang University, kbj0918@hanyang.ac.kr

****Co-Author, College of Economics and Finance, Hanyang University, woneunglee@hanyang.ac.kr