

CFFA와 랜덤 포레스트를 이용한 비상장 건설화학기업의 부실예측

조서연* · 최명수** · 강형구***

〈요 약〉

본 연구는 Comparable firm finding algorithm(CFFA)와 랜덤 포레스트를 활용하여 비상장 건설화학기업의 부실을 예측하였다. CFFA를 통해 기존에는 확인하기 어려웠던 비상장 기업의 시장변수를 측정하여 부족한 비상장 기업의 정보를 보완하였다. 특히 코로나-19의 영향을 중점적으로 분석하고자 2019년부터 2021년까지의 기업 데이터를 활용하여 부실 여부를 예측하였다. 성능평가 결과, CFFA로 측정된 변수를 포함할 경우 검증용 데이터의 모든 평가 지표에서 0.8 이상의 높은 정확도를 보였다. 그에 비해, CFFA로 측정된 변수를 제외할 경우 모든 평가 지표에서 성능 저하가 나타났다. 성능의 차이는 재현율(Recall)에서 가장 크게 나타났다. 따라서 CFFA로 측정된 변수가 부실예측에서 설명력 있는 변수로 적용 가능하며, 부실 기업을 식별하는 능력을 강화함을 확인하였다. 이러한 결과를 통해 본 연구는 기업 및 채권기관이 선제적인 부실관리 대응을 할 수 있는 데 기여할 것으로 기대된다.

주제어 : CFFA, 가치평가, 부실기업 예측, 머신러닝, 랜덤 포레스트

논문접수일 : 2023년 09월 14일 논문수정일 : 2023년 11월 02일 논문게재확정일 : 2023년 11월 15일

* 제1저자, 한양대학교 경영학과 대학원, E-mail: chosoj@hanyanag.ac.kr

** 공동저자, 한양대학교 경영학과 겸임교수, E-mail: kidsjjang@hanyang.ac.kr

*** 교신저자, 한양대학교 파이낸스 경영학과 부교수, E-mail: hyoungkang@hanyang.ac.kr

I. 서 론

코로나-19의 발발 직후 세계 각국은 봉쇄령 및 이동제한 정책을 시행하며 공장 가동률이 대폭 축소되었다. 이로 인해 건설화학 관련 산업이 상당한 피해를 입은 것으로 나타났다. 국내 대기업의 매출 현황을 살펴보면, 건설 및 전자재 업종의 경우 2020년 해외 매출액이 전년 대비 16%나 감소하였으며, 석유화학 업종도 7.9%의 감소를 기록하여 매출 하락의 상위권을 차지하였다.¹⁾ 최근에는 인플레이션, 금리 상승과 같은 경영환경 악화로 기업의 신용 위험과 부실 관리에 대한 관심이 더욱 증가하고 있다.

기업의 부실 위험을 정확히 판별하기 위해서는 경영 부실의 요인을 정확하게 파악하는 것이 중요하다. 그러나 기존의 전통적인 신용평가 및 부실 위험에 관한 연구는 여러 한계점을 가지고 있다. 우선, 부실 위험과 관련된 연구는 대부분 상장 기업에 중점을 두고 이루어져 왔다. 이와 대조적으로 비상장 기업에 대한 연구는 아직까지 부족한 실정이다. 비상장 기업은 공개 정보가 제한적이기 때문에 대출기관이나 투자자 입장에서 이들 기업의 신용 위험을 관리하는 것이 어렵다. 또한, 대기업과는 다른 경영환경에서 운영되므로 재무 구조 측면에서도 차이가 있다. 이에 상장 대기업을 기준으로 한 기존의 리스크 관리 기준은 비상장 기업에 적용되기 어렵다.

또한, 기존의 신용평가 또는 부실 관리와 관련한 기준은 산업적 특성이 고려되지 않아 건설화학 기업을 정확하게 평가하기 어렵다. 한국표준산업분류(대분류)에서는 건설화학 기업을 “일반 제조업”에 분류하고 있지만, 이들 기업은 실제로 건설업과 긴밀한 관련이 있다. 따라서 건설화학 기업을 경영 분석할 때 단순히 “일반 제조업”의 시각만으로 접근하는 것은 적절하지 않다. 화학 제조업과 건설업의 특성을 모두 종합적으로 고려해야 한다. B2B 산업이 주를 이루는 화학제품 제조업은 다른 산업과 비교하여 특히 브랜드의 영향력, 공급자 및 고객 간의 견고한 관계가 기업 가치에 큰 영향을 미친다(Bee, 2018). 한 번 형성된 거래 관계는 쉽게 변화되지 않는 특징을 가지고 있다. 그러므로 기업의 시장 내 위치와 영향력 정도를 파악하는 것이 중요하다. 한편, 건설업은 수주 기반으로 이뤄지는 사업이다. 그 결과 부동산 규제 및 경기상황과 같은 거시 경제적 요인의 영향을 크게 받으며 이로 인한 불확실성과 변동성이 매우 크다(이지혜, 2020; 김구희 외 2인, 2014). 따라서 재무비율 외에도 시장상황에 따른 변동성을 고려하는 것이 매우 중요하다.

본 연구는 Comparable Firm-Finding Algorithm(CFFA)를 사용하여 과거에는 확인하기 어려웠던 비상장 기업의 시장변수를 측정하고, 이를 기반으로 랜덤 포레스트(Random

1) <http://www.worktoday.co.kr/news/articleView.html?idxno=15150>.

Forest)와 같은 머신러닝 기술을 활용하여 기업의 재무 비율 및 시장가치를 모두 고려한 부실예측을 수행할 예정이다. 연구의 목표는 법적인 파산이나 완전한 경영실패를 예측하는 것이 아니라, 경영악화가 예상되는 기업을 사전에 선별하여 기업과 채권기관이 선제적으로 대응할 수 있도록 돕는 데 있다. 앞으로 본 연구는 다음과 같이 진행될 예정이다. II장에서는 기업의 가치평가 및 부실 예측에 관한 선행연구를 검토한다. III장에서는 CFFA와 랜덤 포레스트에 대한 방법론을 설명하고, IV장에서는 CFFA와 랜덤 포레스트의 분석결과에 대해 살펴볼 예정이다. 마지막으로 V장에서는 본 연구의 최종 결과와 의의에 대해 간략히 서술할 예정이다.

II. 선행 연구

1. 가치 평가

일반적으로 현금할인법(Discounted Cash Flow)과 같은 절대가치 평가법은 미래 경영상황이나 경제 전반에 대한 많은 데이터를 필요로 하고, 데이터의 불완전성 또는 부정확성이 있을 경우 신뢰도가 떨어진다는 단점이 존재한다(Steiger, 2010). 이러한 절대가치 평가법은 불확실성이 높은 기업, 예를 들어 건설업이나 벤처기업과 같은 기업에는 부적절하다는 연구결과가 있다(류두진, 박대현, 2021; 김구희 외, 2014).

이에 본 연구에서는 상대가치 평가법 중 유사기업 접근법을 사용하여 비상장 건설화학 기업의 가치평가를 진행할 예정이다. 유사기업분석법(Comparable Company Analysis)은 주로 유사한 상장기업의 정보를 기반으로 비상장 기업의 정보를 추정하는 데 주로 사용된다. 예를 들어, Bowman and Bush(2007)는 유사 상장기업의 평균 베타를 사용하여 480여 개의 미국 비상장 기업의 베타를 추정하였다. 이를 통해 기업 가치 평가에서 유사기업 분석법의 유용성을 입증하였다. 또한, How et al.(2007)는 호주 IPO기업을 대상으로 한 분석에서 타겟기업의 주가수익(Price-to-Earnings; 이하, P/E)과 주가순자산(Price-to-book; 이하, P/B) 배수가 두 유사기업의 평균 P/E와 P/B 배수와 강한 연관성이 있다는 것을 입증하였다. 국내 연구에서도 송혁준, 오웅락(2007)은 현금할인법, 초과이익할인법, 유사기업분석법 중 벤처기업 가치평가에 가장 적합한 평가방법을 찾는 연구를 진행하였다. 각 평가방법의 절대 오차율을 분석한 결과 유사기업분석법이 벤처기업의 가치평가에 가장 적합한 방법임을 입증하였다.

유사기업의 선정기준과 관련하여, Alford(1992)는 미국 기업 사례를 사용하여 유사기업의

선택 기준으로 산업별, 이익성장률 등이 바뀔에 따라 P/E 평가방법의 정확도가 어떻게 달라지는지 비교 분석하였다. 그 결과 산업을 기준으로 유사기업을 선택하는 것이 타겟 기업의 P/E배수 예측에 더 효과적임을 밝혀냈다.

김주은(2022)은 KOSPI 비금융 상장기업의 주가매출비율(Price to sales ratio; 이하, PSR)과 기업성과 간의 관계를 분석했다. PSR이 높은 기업은 더 높은 성장가능성과 가치를 보이며, 하위 기업은 재무구조와 비용 관리 개선이 필요한 것으로 나타났다. 이를 통해 업종에 관계없이 PSR은 기업가치 평가에 유의미한 지표로 작용한다는 것을 확인했다.

기존의 유사기업 분석법은 시장의 현재 상황과 경쟁사들의 동향을 반영하여 상대적 가치를 손쉽게 평가할 수 있다는 장점이 있다. 그러나 기업의 자원과 역량에 대한 정보를 통합하거나, 동종 기업 특성에 대한 가중 평균을 계산하기 어렵다는 한계가 있다. 또한, 기업의 전략적 자원이나 기타 무형자산의 가치를 분리할 수 없다. 본 연구에서 사용하는 CFFA는 회귀분석을 기반으로 유사기업의 가중치를 결정하여 통계적으로 유의미한 결과를 얻을 수 있다. 또한, 빅데이터와 머신러닝 기술을 활용해 다양한 분석이 가능하다. 이를 통해 기업의 전략적 자원 평가, ESG 기여도 식별, 투자 전략 및 시나리오 기획 등 다양한 응용이 가능하다는 장점이 있다(Kang, 2021).

2. 부실 예측

부실예측에 관한 연구는 오랜 기간 동안 다양한 모델과 데이터를 활용되어 진행되어 왔다. 초기에는 Altman(1968)의 다변량 판별분석을 이용한 Z-score모형, Ohlson(1980)의 로짓 모형과 같은 재무적 부실 연구가 이뤄졌다. 이후 부실예측 연구는 재무정보 외에도 시장 정보를 활용하여 부도예측의 정확도와 적시성을 향상시키려는 방향으로 발전해왔다. 대표적으로 Shumway(2001)는 기업의 재무 건전성 지표와 시장 정보를 결합하여 부실 예측력을 높이는 헤저드 모형을 제안했다. 이 모형은 기존의 정적인 환경에서의 재무 데이터만을 고려하는 예측 모델의 한계를 극복하고, 시장 가격 변동 등을 고려하여 예측의 정확성을 높이는 것을 목표로 했다.

더 나아가 Dichev(1998)는 주식수익률과 파산 위험 간의 음(-)의 관계가 있다는 것을 분석했다. 이로써 파산 위험이 높은 기업의 주식이 주식시장의 전반적인 리스크와 관련이 있다는 것을 밝혔다. Minton et al.(2002)은 변동성이 미래 영업현금흐름 및 미래 영업이익과 음(-)의 관계를 가지고 있다고 밝혔다. 이에 변동성이 예측 정확도를 높일 수 있는 유의한 변수로서 활용될 수 있음을 입증하였다. Campbell et al.(2008)은 Shumway의 헤저드 모형을 발전시켜 새로운 변수를 도입하여 예측력을 높였다. 시장 가격이 기업의 전망에 대한 새로운

정보를 더 신속하게 통합하고 회사의 무형자산에 대한 가치를 더 정확하게 반영할 수 있다고 지적했다. 또한, 재정적으로 부실한 주식이 부실 위험이 낮은 주식보다 주식수익률의 표준편차가 더 높다고 지적하였다. Hernandez Tinoco and Wilson(2013)은 회계 데이터, 주식 시장 정보 및 거시 경제 데이터를 종합하여 부실 예측을 수행했다. 그 결과 시장 변수를 포함한 모델이 다른 모델에 비해 더 높은 정확도를 보였고, 거시경제 변수의 중요성은 상대적으로 낮다는 결과를 제시했다.

최근의 부실 예측 연구는 빅데이터와 머신러닝, 인공 지능 기술을 적극적으로 활용하는 방향으로 발전하고 있다. Barboza et al.(2017)은 파산 발생 1년 전 예측을 위해 머신러닝 모델(서포트 벡터 머신, 배깅, 부스팅, 랜덤 포레스트)을 이용해 예측을 진행하고, 그 성능을 전통적인 통계 기법(판별분석, 로지스틱)과 초기 인공지능 모델(인공신경망)과 비교하였다. 머신러닝 모델은 기존 모델에 비해 약 10% 더 높은 정확도를 보였다. 특히 배깅, 부스팅, 랜덤 포레스트 모델이 다른 기술보다 성능이 뛰어났으며, 기존의 Altman의 z-score 외에도 영업이익, 자기자본이익률의 변화, 자산, 매출, 직원 수와 관련된 성장성 지표를 새로운 변수로 포함했을 때 예측의 정확도가 크게 향상됨을 확인하였다. Malakauskas and Lakštutienė(2021)는 머신러닝을 사용하여 1년 동안의 정적인 회계 데이터만 고려하는 것이 아니라, 연체 내역, 자산 및 판매특성의 변화와 같은 시간적 요소를 추가했을 때 예측의 정확도가 높아진다는 결과를 도출했다. 이때 로지스틱, 인공신경망, 랜덤 포레스트 중 랜덤 포레스트가 가장 뛰어난 성능을 보였다고 밝혔다. 국내 연구에서도 김성진, 안현철(2016)은 랜덤 포레스트를 이용해 신용등급을 예측하고, 결과를 전통적인 분류기법인 다중판별분석, 인공신경망, 다분류 SVM 모형과 비교 분석하였다. 그 결과, 이상치와 잡음의 영향이 작은 랜덤 포레스트가 리스크 관리, 신용평가 등 데이터 불균형이 심한 경우가 많은 금융환경에 매우 적합한 알고리즘이라고 밝혔다. 이정환, 조진형(2023)은 머신러닝 기법을 사용하여 이전에는 신용 등급이 없어 분석이 불가능한 비자발적 상장 폐지 기업에 대한 신용평가를 진행하였다. 이를 통해 신용평가 관련 분석을 다양한 기업에 대해 확장할 수 있는 새로운 프레임워크를 제시하였다. 또한, ESG 중 지배구조 정보가 신용 등급 평가를 위한 주요 정보가 될 수 있다고 밝혔다.

Ⅲ. 자료와 연구방법론

1. 자료

본 연구는 건설소재 화학제품 제조업을 영위하는 외부감사대상법인 중 국내 유가증권

시장에 상장되지 않은 비상장 기업을 대상으로 한다. 건설소재 화학 산업이란 건축작업 시 시멘트, 콘크리트 또는 기타 건축 자재와 함께 사용되는 화학 약제를 제조하는 분야를 의미한다.²⁾ 따라서 연구대상의 범위를 다음의 통계청 한국표준산업분류10차(세세분류) 기준에 속하는 기업들로 제한하였다: 석유화학계 기초화학물질 제조업, 석탄화학계 화합물 및 기타 기초 유기 화학물질 제조업, 기타 기초 무기 화학물질 제조업, 일반용 도료 및 관련제품 제조업, 계면활성제 제조업, 그 외 기타 분류 안 된 화학제품 제조업.

데이터의 누락이 있는 기업, 건설화학과 관련 없는 기업은 모두 제외되었다. 이로써 최종적으로 120개 기업(상장 34개사 및 비상장 86개사)이 선정되었다. 연구의 주요 대상은 비상장 기업 86개사로, 상장기업 34개사는 비상장 기업의 분석을 지원하기 위해 활용될 것이다. 연구에 사용된 자료의 기간은 2019년부터 2021년까지이다. 각 기업의 재무데이터는 FnGuide의 데이터를 활용하였으며, 일부 결측치의 경우 직접 계산하여 보완하였다.

2. 연구방법

1) Comparable Firm Finding Algorithm(CFFA)

Kang(2021)는 Comparable firm finding algorithm(CFFA) 모델을 처음으로 제안하였다. 이 모델은 전략적 자원, 회귀 및 포트폴리오 이론을 기반으로 한 직관적인 가치 평가 프레임워크이다. CFFA는 유사기업 분석법(Comparable Company Analysis; 이하, CCA) 방식을 일반화하여 더 폭넓은 적용과 변형이 가능하도록 한다. CCA를 포함한 기존의 가치평가 모델은 기업의 자원과 역량에 대한 정보를 통합하거나, 동종 기업 특성에 대한 가중 평균을 계산하기 어렵다는 한계가 있다. 또한, 기업의 전략적 자원이나 기타 무형자산의 가치를 분리할 수 없다. 하지만 무형 자산 및 기업의 비재무적 데이터의 중요성이 높아지는 시대에 기존의 가치평가 방식은 적합성과 실용성이 떨어진다는 문제점이 있다.

CFFA는 회귀분석 방법을 사용하여 이러한 한계를 모두 극복할 수 있다. CFFA는 회귀분석 방식을 통해 타겟 기업과 유사 기업 간의 유사성 정도를 정량화하고 회귀계수를 가중치로 할당한다. 회귀분석을 기반으로 타겟기업에 대한 유사기업의 가중치가 결정되므로 통계적 유의성을 얻을 수 있다. 만약 타겟 기업과 유사기업이 서로 유사하지 않을 경우에는 가중치가 거의 할당되지 않는다. 따라서 타겟 기업에 대한 보다 강력한 피어 그룹을 설정할 수 있다는 장점이 있다. 또한, CFFA는 1, 2개의 재무비율만을 사용하는 CCA와는 달리 많은 데이터를

2) https://www.designingbuildings.co.uk/wiki/Construction_chemicals.

활용할 수 있으며, 빅데이터 분석이나 머신 러닝과 같은 기술을 활용해 통계 기술을 확장할 수 있는 유연성을 갖는다. 이를 통해 기업의 전략적 자원과 무형자산에 대한 가치평가, ESG 기여도 식별, 투자 전략 및 시나리오 기획에 효과적으로 적용할 수 있다(Kang, 2021).

본 연구에서는 CFFA를 활용하여 기존에 관측하기 어려웠던 비상장 기업의 시가총액과 변동성 지표를 측정하고, 이를 활용해 기업의 부실예측을 진행할 예정이다. 많은 선행 연구들과 마찬가지로, 기업의 시장가치는 부실예측에 있어서 유의미한 영향을 미칠 것으로 예상된다(Campbell et al., 2008; Hernandez Tinoco and Wilson, 2013; Shumway, 2001). Comparable firm finding algorithm(CFFA)를 활용하여 시가총액을 계산함으로써 비상장 기업의 시장 내 위치와 브랜드 영향력 등을 파악할 계획이다. 이를 통해 기업이 부실위험을 얼마나 견딜 수 있는지를 평가할 수 있을 것이다. 더불어 시장 상황에 따른 변동성 지표를 측정하여 재무제표만으로 파악하기 어려웠던 시장 위험에 대한 취약성을 평가할 수 있을 전망이다. 이와 같은 추가적인 시장 변수 고려를 통해 기업의 부실예측 정확도를 높일 수 있을 것으로 기대된다.

CFFA 적용 과정은 다음과 같다. 우선, 연구대상인 비상장 기업을 타겟기업으로 선정하고 각 타겟기업과 유사한 상장기업들을 찾아 피어 그룹을 구성한다. 이후 가치평가에 사용할 각 기업들의 속성 정보를 수집한다. 속성에는 당기순이익률, ROA와 같은 재무적 요소 뿐만 아니라 텍스트 정보나 감성정보 등도 포함될 수 있다. 단, 후속의 회귀분석을 위해 속성의 수는 유사기업의 수보다 많아야 한다($T < N$). 수집한 속성 정보는 기업별로 하나의 열로 묶어 각각의 벡터로 구성한다. 주요변수는 다음과 같이 나타낼 수 있다.:

N : 기업별 속성의 수

T : 유사기업의 수($T < N$)

\vec{y} : 타겟기업의 속성에 대해 N 행으로 구성된 열 벡터

\vec{x}_i : 유사기업 i 의 속성에 대해 N 행으로 구성된 열 벡터

x : 피어 그룹의 속성에 대한 $N \times T$ 행렬

$$x \equiv (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_T)$$

타겟기업에 대한 벡터 \vec{y} 는 종속변수에 해당한다. 유사기업에 대한 벡터를 \vec{x}_i 로 정의할 때, 이러한 유사기업들의 집합인 피어 그룹을 행렬 x 로 둘 수 있다. 이 피어그룹 전체가 독립변수에 해당한다. 피어 그룹의 속성을 타겟 기업과 연관시키는 방법은 다음과 같은 회귀 접근법으로 가능하다.

$$\vec{y} = x\vec{\beta} + \vec{\epsilon} \quad (1)$$

이때 $\vec{\beta}$ 는 합성 유사기업을 구성하기 위한 포트폴리오 가중치로 볼 수 있다. $x\vec{\beta}$ 는 새롭게 구축된 합성 타겟기업의 속성이라 할 수 있다. $\vec{\epsilon}$ 는 실제 타겟기업과 합성 타겟기업의 속성 사이의 오류를 나타낸다. 단, 본 연구는 타겟기업과 유사기업 사이의 관계를 회귀 분석하고자 함이 목적이므로 상수항은 제거된다. 타겟기업의 합성버전을 추정하기 위해 OLS(Ordinary least squares), GLS(Generalized least squares), Lasso 등의 회귀분석 방법을 사용할 수 있다. 일반적인 OLS 회귀분석을 진행한다고 가정한다면 타겟기업에 대한 유사기업의 포트폴리오 가중치는 다음과 같이 구할 수 있다.

$$\beta_{OLS} = (x'x)^{-1}x'y \quad (2)$$

위와 같은 과정을 통해 각 타겟 기업에 대한 유사 기업의 가중치를 도출한 후, 연구자는 이를 활용하여 추정하고자 하는 값을 계산할 수 있다. 예를 들어, 타겟기업의 주가매출비율(price-to-sales ratio, PSR)을 구하고자 한다면 다음과 같은 방법을 사용할 수 있다. 유사기업의 실제 주가매출비율(PSR)과 유사기업의 포트폴리오 가중치를 곱한 후, 이들의 전체 합으로 타겟기업의 주가매출비율을 추정할 수 있다. 이를 식으로 표현하면 다음과 같다.

$$PSR_{Target} = (PSR_1, PSR_2, \dots, PSR_T) \cdot \beta_{OLS} \quad (3)$$

2) 랜덤 포레스트(Random Forest)

랜덤 포레스트(Random Forest; 이하, RF)는 Breiman(2001)이 처음 제안한 앙상블 기반 머신러닝 모델로 분류와 회귀분석에 모두 사용 가능하다. RF는 데이터가 입력된 단일 의사결정 나무(Decision Tree)를 숲(Forest)으로 확장하여 훈련하고, 그 결과를 종합하여 예측 결과를 얻는 방식이다. 이때 각 의사결정 나무의 학습 데이터는 무작위로 선택되며, 각 분기에서 사용될 특성(변수)도 무작위로 선택된다. 또한, 모델의 학습 데이터는 부트스트랩 샘플링을 통해 생성되어 더욱 다양한 학습 데이터를 활용할 수 있다. 부트스트랩은 훈련 데이터에서 표본 데이터 세트를 추출할 때 중복을 허용하여 원 데이터 세트와 같은 크기의 데이터 세트를 만드는 과정이다. RF는 이러한 무작위성을 통해 다른 머신러닝 알고리즘에

비해 과적합의 영향을 줄일 수 있으며, 잡음이나 이상치의 영향을 크게 받지 않는다는 장점을 갖는다. 따라서 RF는 빈도가 불균형한 이항분류 예측에 있어 가장 우수한 예측력을 보인다. 이는 신용위험 또는 부실관리와 같이 데이터가 불균형한 경우에 매우 효과적으로 적용될 수 있다는 점을 시사한다(김성진, 안현철, 2016).

RF를 이용한 부실예측 과정은 다음과 같다. 먼저, 부트스트랩을 통해 분석 데이터에서 임의의 부분 집합을 생성한다. 그 다음, 각 부분 집합에 대해 의사결정 나무를 생성한다. 생성된 다수의 의사결정나무로부터 개별 예측결과를 얻으면 각 결정 트리들이 예측한 결과를 다수결 투표(Majority vote)하여 최다 표를 받은 범주를 채택하여 분류한다. $\hat{C}_b(x)$ 를 b번째 부트스트랩 훈련 세트의 예측치라고 가정할 때, 결과를 다음 수식과 같이 나타낼 수 있다(이정환, 조진형, 2023).

$$\hat{C}_{rf}^B(x) = \text{majority vote} \{ \hat{C}_b(x) \}_1^B \quad (4)$$

* B: 부트스트랩 개수(b: 1부터 B까지)

IV. 실증분석

1. CFFA를 이용한 시장변수 측정

이제 CFFA를 이용하여 타겟기업의 시장변수 즉, 시가총액과 주식수익률의 변동성을 측정한다. 가치평가를 위해 상장 기업과 비상장 기업 모두의 속성 정보가 활용된다. 이에 사용되는 속성은 기업의 수익성(매출총이익률, 영업이익률, 당기순이익률, ROE, ROA), 성장성(매출액증가율, 총자산증가율, 영업이익증가율, 자기자본증가율, 무형자산구성비율, 연구개발비율), 활동성(총자산회전율, 매입채무회전율, 매출채권회전율, 재고자산회전율, 판매관리비/매출액), 안정성(부채비율, 유동비율, 자기자본비율) 등의 재무지표 및 기타 정보(기업연령, 기업규모)로 구성된다. 특히, R&D 투자, 무형자산 등이 기업가치에 영향을 미친다는 선행 연구 결과들을 참고하여 이를 성장성 지표에 포함시켰다(전성일, 이병원, 2003). 변수들의 기초 통계량 정보는 <표 1>에 제시하였다. 분석에 사용되는 변수는 연도별로 구분하여 별도의 속성으로 취급하였다. <표 1>에서 약어로 표기된 속성의 정확한 변수명은 별도로 표 하단에 기재하였다. 각 속성별 측정값 수(N), 평균, 표준편차, 중앙값, 최소값, 최대값을 표기하였다. 데이터 분포가 정규분포를 따르지 않고 한쪽으로 쏠린 속성 정보가 많아 다음과 같이 변환 작업을 수행하였다. Expense, CR, ACPTO, RnD, ITO, NPM 데이터는

로그 변환하였다. OPM은 지수 변환하였으며, ROE, AGR은 제곱근 변환을 진행했다. Debt, OPM, Intangible은 박스-콕스(Box-Cox) 변환을 시켰다. 이 과정에서 값이 0이 나오거나 음의 왜도(Negative skew)를 보이는 경우에는 상수를 더하거나 양의 왜도(Positive skew)로 변환한 후에 변환 작업을 진행했다.

<표 1> 기초 통계량

<표 1>은 회귀분석에 사용된 각 기업의 속성에 대한 기초 통계량이다. 분석에 사용되는 변수는 연도별로 구분하여 별도의 속성으로 취급하였다. 약어로 표기된 속성명은 표 하단에 별도 표기하였다. 속성별 측정값 수(N), 평균, 표준편차, 중앙값, 최소값, 최대값을 차례대로 표시하였다. 자료의 기간은 2019년부터 2021년까지이다. 분포가 정규분포를 따르지 않고 한쪽으로 쏠린 속성 정보가 많아 다음과 같이 변환 작업을 진행하였다. Expense(판매관리비/매출액 비율), CR(유동비율), ACPTO(매입채무회전율), RnD(연구개발비율), ITO(재고자산회전율), NPM(당기순이익률) 데이터는 로그 변환하였다. OPM(영업이익률)은 지수 변환하였으며, ROE, AGR(총자산증가율)은 제곱근 변환을 진행했다. Debt(부채비율), OPM(영업이익증가율), Intangible(무형자산구성비율)은 박스-콕스(Box-Cox) 변환을 시켰다. 이 과정에서 값이 0이 나오거나 음의 왜도(Negative skew)를 보이는 경우에는 상수를 더하거나 양의 왜도(Positive skew)로 변환한 후에 변환 작업을 진행했다.

	N	Mean	Stdev	Median	MIN	MAX
SIZE_19	120	18.5367	1.8619	17.87	16.29	24.25
SIZE_20	120	18.5664	1.8709	17.94	16.30	24.45
SIZE_21	120	18.6717	1.9412	18.00	16.09	24.66
AGE_19	120	29.2583	15.6337	29.00	2.00	73.00
AGE_20	120	30.2583	15.6337	30.00	3.00	74.00
AGE_21	120	31.2583	15.6337	31.00	4.00	75.00
Intangible_19	120	0.3380	0.2910	0.29	0.00	0.88
Intangible_20	120	0.3346	0.2962	0.29	0.00	0.88
Intangible_21	120	0.3221	0.2985	0.26	0.00	0.88
RnD_19	120	0.4012	0.5490	0.07	0.00	2.10
RnD_20	120	0.4158	0.5573	0.06	0.00	2.22
RnD_21	120	0.3593	0.4930	0.06	0.00	2.26
Expense_19	120	2.3075	0.7449	2.43	-0.27	4.33
Expense_20	120	2.3881	0.7378	2.45	0.02	4.30
Expense_21	120	2.2973	0.8334	2.31	-0.36	5.38
GPM_19	120	18.6384	10.8704	17.10	1.54	82.38
GPM_20	120	19.7325	10.9645	18.13	3.21	81.89
GPM_21	120	19.1655	11.9094	16.64	-8.38	81.88
OPM_19	120	7.8424	0.5277	7.79	5.97	9.88
OPM_20	120	7.8396	0.5976	7.84	5.42	9.77
OPM_21	120	7.8490	1.0374	7.84	1.29	11.80
NPM_19	120	3.4195	0.3232	3.46	1.32	4.80
NPM_20	120	3.4237	0.4208	3.44	0.46	5.42
NPM_21	120	3.3626	0.5674	3.41	0.00	6.15
SGR_19	120	-3.0202	14.7252	-3.52	-58.85	61.24
SGR_20	120	-3.6195	17.7978	-2.64	-82.07	86.93
SGR_21	120	24.4864	27.6930	21.43	-96.81	151.42

<표 1> 기초 통계량(계속)

	N	Mean	Stdev	Median	MIN	MAX
AGR_19	120	7.6456	1.2085	7.55	2.81	15.13
AGR_20	120	7.5283	0.8376	7.61	4.38	9.65
AGR_21	120	8.0605	1.2488	7.92	4.05	13.70
OPGR_19	120	39.8400	6.9625	39.84	0.00	100.15
OPGR_20	120	39.5976	5.6291	40.24	12.12	63.21
OPGR_21	120	40.3430	5.6830	40.26	18.00	70.52
OEGR_19	120	2.8311	88.4028	4.61	-906.42	205.35
OEGR_20	120	0.8482	26.1408	4.68	-217.33	35.39
OEGR_21	120	23.4049	149.3562	8.73	-92.86	1,625.03
OE_19	120	56.6795	20.6150	55.63	9.47	92.88
OE_20	120	55.5623	22.0048	57.02	-31.84	93.24
OE_21	120	53.6640	22.9928	54.33	-67.02	91.94
ROE_19	120	28.9533	1.0432	29.04	18.02	30.60
ROE_20	120	29.2256	2.6631	29.03	25.47	57.82
ROE_21	120	29.0367	0.5000	29.06	25.72	30.68
ROA_19	120	4.5608	8.6852	4.55	-59.21	23.15
ROA_20	120	3.9918	9.2497	4.03	-55.89	24.71
ROA_21	120	6.4883	10.5140	5.92	-31.97	60.05
Debt_19	120	11.2084	0.4153	11.11	10.80	13.48
Debt_20	120	11.2563	1.0818	11.07	7.91	21.77
Debt_21	120	11.2910	1.0037	11.12	9.40	21.16
CR_19	120	5.1663	0.7008	5.03	3.79	7.15
CR_20	120	5.1412	0.7468	5.06	3.06	6.99
CR_21	120	5.0953	0.7339	5.03	1.90	6.91
ATO_19	120	1.1045	0.4764	1.01	0.28	2.86
ATO_20	120	1.0021	0.4203	0.95	0.17	3.13
ATO_21	120	1.1453	0.4822	1.10	0.03	3.10
ACPTO_19	120	2.5950	0.7996	2.66	0.81	4.89
ACPTO_20	120	2.5198	0.7624	2.58	0.74	4.89
ACPTO_21	120	2.5389	0.7442	2.55	1.15	5.43
ACRTO_19	120	6.8464	3.8600	6.34	1.69	30.89
ACRTO_20	120	6.6581	3.5635	6.17	1.32	24.72
ACRTO_21	120	7.4699	3.9347	6.82	0.52	25.88
ITO_19	120	2.4158	0.5417	2.37	1.25	5.05
ITO_20	120	2.3725	0.5655	2.36	1.20	5.54
ITO_21	120	2.3443	0.5730	2.31	0.00	4.15
* <i>SIZE</i> : 기업규모						
* <i>RnD</i> : 연구개발비용						
* <i>OPM</i> : 영업이익률(%)						
* <i>AGR</i> : 총자산증가율(%)						
* <i>OE</i> : 자기자본비용(%)						
* <i>Debt</i> : 부채비용(%)						
* <i>ACPTO</i> : 매입채무회전율						
* <i>AGE</i> : 기업연령						
* <i>Expense</i> : 판매관리비/매출액						
* <i>NPM</i> : 당기순이익률(%)						
* <i>OPGR</i> : 영업이익증가율(%)						
* <i>ROE</i> : ROE						
* <i>CR</i> : 유동비용(%)						
* <i>ACRTO</i> : 매출채권회전율						
* <i>Intangible</i> : 무형자산구성비용						
* <i>GPM</i> : 매출총이익률(%)						
* <i>SGR</i> : 매출액증가율(%)						
* <i>OEGR</i> : 자기자본증가율(%)						
* <i>ROA</i> : ROA						
* <i>ATO</i> : 총자산회전율						
* <i>ITO</i> : 재고자산회전율						

건설화학 산업은 최종 제품의 수요시장 상황에 따라 사업 성과가 크게 영향 받기 때문에, 각 기업의 주요 사업을 기준으로 유사기업을 선정하였다. 유사기업은 통계청 한국표준산업 분류10차(세세분류) 기준을 따라 주요 사업 구분 기준을 설정하였으며, 이를 바탕으로 유사기업은 6개의 그룹으로 분류되었다. 이후 타겟 기업을 종속변수 \vec{y} 로, 해당 타겟 기업의 유사기업들을 독립변수 \vec{x}_i 로 두고 분석을 진행하였다. 이때 \vec{x}_i , \vec{y} 는 모두 각 기업의 속성정보가 하나의 열로 묶인 벡터로 볼 수 있다. 유사기업 간의 강한 상관관계로 인한 다중 공선성(multicollinearity) 문제를 고려하여 라쏘 회귀분석(Lasso regression)을 사용하였다. 라쏘 회귀는 계수의 절대값을 줄이는 L1규제를 사용하여 중요하지 않은 변수에 해당하는 가중치를 0으로 만드는 방식이다(Tibshirani, 1996). 과적합을 방지하기 위해서 잔차제곱합에 패널티 함수 $\alpha \sum_{j=1}^p |\beta_j|$ 를 추가하여 다음 식을 가장 작게 하는 β 를 추정한다. 라쏘 회귀를 통해 추정된 타겟 기업의 속성에 매치되는 유사기업의 포트폴리오 가중치는 다음과 같은 식으로 나타낼 수 있다(권누리 외 2인, 2019; 이재득, 2021).

$$\hat{\beta}_{Lasso} = \arg \min_{\beta} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{i,j} \beta_j)^2 + \alpha \sum_{j=1}^p |\beta_j| \quad (5)$$

여기서 α 는 규제 파라미터로, 이 값이 높아지면 타겟기업과 관련이 적은 유사기업의 추정 계수 β 는 0으로 수렴한다. 반대로 α 값이 낮아질수록 패널티 효과 감소로 인해 β 가 커지고 OLS 선형회귀모형에 가까워진다.

한편, 본 연구에서 \vec{x} , \vec{y} , β , 는 모두 벡터이므로 라쏘 회귀식을 본 연구에 맞게 수정할 필요가 있다. $\beta_0 = 0$ 으로 제거되며, 피어그룹 벡터 $x \equiv (x_1^-, x_2^-, \dots, x_T^-)$, 회귀계수 벡터 $\beta \equiv (\beta_1, \beta_2, \dots, \beta_p)$ 일 때 위 식은 다음과 같이 변경될 수 있다.

$$\beta_{Lasso} = \min_{\beta} \|\vec{y} - x\beta\|_2^2 - \alpha \|\beta\|_1 \quad (6)$$

여기서 $\|\beta\|_1$ 와 $\|\vec{y} - x\beta\|_2^2$ 은 다음을 의미한다.

$$\begin{aligned} \|\beta\|_1 &= |\beta_1| + |\beta_2| + \dots + |\beta_p|: \text{계수의 절대값의 합} \\ \|\vec{y} - x\beta\|_2^2 &= (\vec{y} - x\beta)'(\vec{y} - x\beta): \text{잔차의 제곱합} \end{aligned}$$

각 타겟 기업에 대한 회귀분석 결과는 <표 2>에 표기하였다. 종속변수는 비상장 기업인 타겟기업이며, 각 패널 첫 번째 행에 기업명을 표기하였다. 독립변수는 상장 기업인 유사기업으로, 각 패널 첫 열에 기업명을 표기하였다. 추가적으로 각 패널 상단에는 alpha

값을 표기하였다. 그러나 일부 기업의 경우 유의미한 R^2 값을 얻기 위해 alpha 값을 조정하였다. 이 경우 기업명과 함께 alpha 값을 괄호로 별도로 표기하였다. 또한, Panel D의 SAMGW, Panel E의 SIKA의 경우에는 시가총액 추정 과정에서 음수가 나오지 않도록 회귀분석 시 양수인 계수만 출력되도록 scikit-learn 라이브러리에서 제공하는 제약조건을 적용하여 계산하였다.

<표 2> 기업별 가중치 확인

<표 2>는 각 타겟 기업에 대한 회귀분석 결과를 나타낸다. 회귀분석 방법은 다중 공선성 문제를 고려하여 라쏘 회귀(Lasso regression)를 사용하였다. 식에서 벡터 x 는 피어 그룹을 의미하며, 벡터 β 는 합성 유사기업을 구성하기 위한 포트폴리오 가중치를 의미한다. 이때 타겟기업과 관련이 적은 유사기업의 추정 계수 β 는 0으로 수렴한다. 종속변수는 비상장 기업인 타겟기업이며, 각 패널 첫 번째 행에 기업명을 표기하였다. 독립변수는 상장 기업인 유사기업으로, 각 패널 첫 열에 기업명을 표기하였다. 추가적으로 각 패널 상단에는 alpha 값을 표기하였다. 그러나 일부 기업의 경우 유의미한 R^2 값을 얻기 위해 alpha 값을 조정하였다. 이 경우 기업명과 함께 alpha 값을 괄호로 별도로 표기하였다. 또한, Panel D의 SAMGW, Panel E의 SIKA의 경우에는 시가총액 추정 과정에서 음수가 나오지 않도록 회귀분석 시 양수인 계수만 출력되도록 scikit-learn 라이브러리에서 제공하는 제약조건을 적용하여 계산하였다.

$$\beta_{Lasso} = \min_{\beta} \|\vec{y} - x\beta\|_2^2 + \alpha \|\beta\|_1$$

(6)

Panel A: 석유화학계 기초 화학물질 제조업(alpha=30.88)

Independent variable	SNF	YNCC	LTIN-EOS	BASF(alpha=80)	HTE	LTMCC	EIL	SKGEO	SKADV	SKPUC
LG	0.2584	0.0306	0	0.1616	0	0	0.1218	0	0.0382	0.0603
HH	0.0973	0	0	0	0	0	0	0	0	0
LOTTE	0	0.5933	0	0	0.5598	0	0.5166	0.4244	0.1845	0
DH	0.5005	0	0	0.5642	0	0.6053	0	0	0.2321	0.5795
ISU	0	0	0.0043	0	0	0	0	0	0	0
AK	0	0.1152	0.8884	0.0418	0.1998	0.2930	0	0.2856	0.2882	0
KUKDO	0	0	0.2847	0	0	0	0	0	0	0
R^2	0.8554	0.8215	0.7234	0.8555	0.8849	0.8688	0.8230	0.8436	0.8313	0.6586
$Adj.R^2$	0.8370	0.7988	0.6882	0.8371	0.8703	0.8521	0.8005	0.8237	0.8098	0.6152

Panel B: 석탄화학계 화합물 및 기타 기초 유기 화학물질 제조업(alpha=100)

Independent variable	LXMMMA	KHMTS	PNB	SAM-NAM	SAKM	SHINWN(alpha=130)	DOW	HAEIN
KYANG	0	0	0.0214	0	0	0	0	0
LTFINE	0.5165	0.5888	0.3962	0.1827	0.4358	0.0794	0.4799	0.3492
MWSC	0.2422	0.2662	0.4512	0	0.2508	0.0565	0	0
KG	0	0	0	0	0	0	0	0
KPX	0.0383	0	0	0.4274	0	0.6584	0.0215	0.2636
HKAC	0	0	0	0	0	0	0	0
R^2	0.8932	0.8869	0.6740	0.7660	0.8286	0.8998	0.7445	0.7144
$Adj.R^2$	0.8817	0.8748	0.6391	0.7409	0.8103	0.8891	0.7171	0.6838

<표 2> 기업별 가중치 확인(계속)

Panel B: 석탄화학계 화합물 및 기타 기초 유기 화학물질 제조업(alpha=100)

Independent variable	KUMJNG	ALMIN	KUDHIN	SINGW	EMTC	WZHT
KYANG	0	0.1722	0	0	0	0
LTFINE	0.3580	0.0255	0.2630	0.3329	0.1242	0.0357
MWSC	0.1539	0.2318	0.1372	0	0.4480	0.1968
KG	0.0661	0	0	0.0523	0	0
KPX	0	0	0	0.1525	0	0.2554
HKAC	0	0	0.2251	0	0	0
R^2	0.7717	0.5372	0.8304	0.5423	0.6022	0.6836
$Adj.R^2$	0.7472	0.4877	0.8122	0.4933	0.5596	0.6497

Panel C: 계면활성제 제조업(alpha=12.06)

Independent variable	DDC (alpha=70)	SNZN (alpha=110)	SNKL (alpha=100)	JWHIN	AEZIS	HWSNG
GREEN	0	0	0	0.4412	0.6932	0
DNAM	0.0997	0.2242	0.0882	0.1390	0.0438	0.2513
HNONG	0.7524	0.5165	0.5459	0.0821	0	0
MIWON	0	0	0	0	0.0538	0.4838
R^2	0.8743	0.8932	0.8658	0.6648	0.8759	0.4191
$Adj.R^2$	0.8657	0.8858	0.8565	0.6417	0.8673	0.3790

Panel D: 기타 분류 안 된 화학제품 제조업(alpha=100)

Independent variable	YWC	SILK	ECON (alpha=150)	HWIN	AKEMA	DHPAC (alpha=150)	SUNGU (alpha=130)	DUBON
COSMO	0	0	0	0.0846	0.0494	0	0	0.6015
TKGHU	0	0.2247	0	0	0.0211	0	0.1885	0
SW	0.2152	0.4359	0	0.4423	0	0	0	0
CTNX	0	0	0	0	0	0	0	0.1930
CTRS	0	0	0	0	0	0	0	0
LION	0.2242	0	0	0.0363	0.4630	0	0	0
KDC	0.0948	0	0.7674	0	0	0.8164	0.5158	0
R^2	0.7111	0.6670	0.8856	0.7884	0.5394	0.9001	0.8778	0.2655
$Adj.R^2$	0.6744	0.6247	0.8710	0.7614	0.4808	0.8874	0.8622	0.1720

<표 2> 기업별 가중치 확인(계속)

Panel D: 기타 분류 안 된 화학제품 제조업(alpha=100)

Independent variable	IPSI	NOFMT	NBCNA	SKHS	SAMJI	SAMGW (alpha=0.001)	CMSTAR	ESPC
COSMO	0	0	0	0	0.0977	0.2923	0	0
TKGHU	0	0	0	0	0.2419	0	0.0446	0.0263
SW	0.1599	0	0	0.0434	0	0	0	0
CTNX	0	0	0	0	0	0.5015	0	0
CTRS	0	0	0	0	0	0	0	0.1001
LION	0	0	0	0.1917	0.5125	0	0	0
KDC	0.6693	0.8940	0.8272	0.1451	0	0	0.7638	0.4676
R^2	0.8541	0.8255	0.8631	0.4926	0.8884	0.2236	0.8962	0.7538
$Adj.R^2$	0.8355	0.8033	0.8457	0.4280	0.8742	0.1248	0.8830	0.7225

Panel D: 기타 분류 안 된 화학제품 제조업(alpha=100)

Independent variable	DHPOLI	DYCNE	FLOW	SILITEC	ACNANO	SAEHAN	LUVC	ESC
COSMO	0	0.1745	0	0.0002	0	0	0	0
TKGHU	0.0335	0.2959	0.2275	0.4431	0	0	0.0840	0.2582
SW	0	0	0	0.1079	0	0	0.3211	0
CTNX	0	0	0	0.0592	0.1762	0	0	0
CTRS	0	0	0	0	0.0679	0	0	0
LION	0.0673	0	0	0	0	0.1797	0	0.1165
KDC	0.5997	0	0.2429	0	0.4971	0.4943	0.0693	0
R^2	0.8964	0.4845	0.4392	0.6670	0.6727	0.8506	0.5655	0.4719
$Adj.R^2$	0.8832	0.4188	0.3678	0.6246	0.6310	0.8316	0.5103	0.4047

Panel D: 기타 분류 안 된 화학제품 제조업(alpha=100)

Independent variable	NCH	HAMIL (alpha=110)	BNB	CAS (alpha=0.001)
COSMO	0	0	0.0158	-2.0960
TKGHU	0	0.2600	0	-3.0068
SW	0.1836	0	0	6.6223
CTNX	0	0	0.2508	-3.8137
CTRS	0	0	0.0990	6.6645
LION	0	0	0.1510	-2.1332
KDC	0.5060	0.4270	0	-1.5579
R^2	0.4220	0.8841	0.5484	0.3657
$Adj.R^2$	0.3485	0.8693	0.4909	0.2850

<표 2> 기업별 가중치 확인(계속)

Panel E: 일반용 도료 및 관련 제품 제조업(alpha=0.22)

Independent variable	SDP	BSP	KYDO (alpha=80)	IPK	JUNBS	NYC	PNC	HANJIN
KCC	-0.1038	-0.2142	0	-0.2298	-0.0828	0.0179	0.1566	-0.2763
SAMHWA	0.0769	0.1525	0.0880	0.2479	0.1003	0.1097	0.0510	0.3905
NORUP	0.7590	0.0252	0.0762	0.1710	0.6464	0.4438	0.5189	1.0730
GNJB	-0.0330	-0.6143	0.4972	-0.2164	-0.1295	0.0570	-0.1385	-0.1063
JGP	0.4733	1.5125	0.0019	0.9084	0.3124	0.1723	0.1705	-0.2326
R^2	0.8087	0.7046	0.8686	0.8488	0.7853	0.8634	0.6276	0.2593
$Adj.R^2$	0.7920	0.6787	0.8571	0.8355	0.7664	0.8514	0.5949	0.1943

Panel E: 일반용 도료 및 관련 제품 제조업(alpha=74)

Independent variable	NRCOIL	DONGS	BYC	JUNGS	NRC- HEM	AKZO	DNC	CHUNIL
KCC	0	0	0.0679	0	0	0	0.1544	0
SAMHWA	0	0.3722	0.0815	0	0	0	0	0
NORUP	0.5018	0.1409	0.0035	0.2824	0.0361	0.5058	0.4365	0.2706
GNJB	0.2089	0.2108	0.0362	0.0005	0.3462	0	0.1826	0.3019
JGP	0	0	0.3902	0.4965	0.1770	0.2201	0	0.2069
R^2	0.8487	0.8214	0.7833	0.8532	0.8139	0.8145	0.8076	0.8751
$Adj.R^2$	0.8354	0.8057	0.7642	0.8403	0.7976	0.7982	0.7907	0.8642

Panel E: 일반용 도료 및 관련 제품 제조업(alpha=74)

Independent variable	SAMJING	DHP	BGW	HSCHEM	HEPS	ILSIN	SIKA (alpha=0.001)
KCC	0	0	0	0	0.0216	0	0
SAMHWA	0.3974	0	0.1775	0	0	0	0.2485
NORUP	0.1888	0.2389	0.4188	0.5014	0.2589	0.6785	0.4959
GNJB	0	0.2716	0.1285	0.2918	0.1291	0.0877	0
JGP	0	0.4132	0.0652	0	0.1694	0	0
R^2	0.7977	0.8464	0.8452	0.8757	0.7742	0.8602	0.2626
$Adj.R^2$	0.7800	0.8329	0.8316	0.8648	0.7544	0.8479	0.1979

Panel F: 기타 기초 무기 화학물질 제조업(alpha=2.32)

Independent variable	SCT (alpha=78)	OMIA	GORI	SVSL	SGCHEM (alpha=78)
HANSOL	0.4358	10.2740	0.1146	-2.2889	0.2036
OCI	0	4.0002	-0.1657	6.1275	0
HSUNG	0	0.3688	0	1.6771	0
TK	0.3297	-20.6065	1.2327	-2.3072	0.7940
HANIL	0.2788	7.0650	-0.2488	-1.9607	0
R^2	0.8829	0.1622	0.8219	0.2584	0.8024
$Adj.R^2$	0.8726	0.0887	0.8063	0.1933	0.7850

이를 바탕으로 첫 번째 시장변수인 시가총액을 측정하는 과정은 다음과 같다. 먼저 유사기업의 실제 PSR에 <표 2>에서 계산한 유사기업의 가중치를 곱한 후, 이들의 총합을 계산하여 합성 타겟기업의 PSR을 구한다. 이후, 합성 타겟기업의 PSR에 실제 타겟기업의 매출액을 곱하여 시가총액을 계산한다. 이 과정을 수식으로 나타내면 아래와 같다.

$$PSR_{Target} = (PSR_1, PSR_2, \dots, PSR_p) \cdot \beta_{Lasso} \quad (7)$$

$$PSR_{Target} \left(= \frac{\text{시가총액}}{\text{매출액}} \right) \times \text{매출액}_{Target} = \text{시가총액}_{Target} \quad (8)$$

합성 타겟기업의 PSR 및 시가총액 측정 결과는 <표 3>에 표시하였다. (1)~(3)열은 차례대로 2019년, 2020년, 2021년의 PSR을 나타낸다. (4)~(6)열은 2019년, 2020년, 2021년에 해당하는 시가총액 계산 결과이다.

<표 3> PSR 및 시가총액 측정

<표 3>는 합성 타겟기업의 PSR 및 시가총액 측정 결과를 나타낸다. 첫 번째 시장변수인 시가총액을 계산하기 위하여 먼저 합성 타겟기업의 주가매출비율(PSR)을 측정하였다. 이는 유사기업의 실제 PSR에 <표 2>에서 구한 유사기업의 포트폴리오 가중치를 곱한 후, 이에 대한 전체 합으로 계산된다. 이후 합성 타겟기업의 PSR에 타겟기업의 실제 매출액을 곱하여 시가총액을 측정한다. (1)~(3)열은 차례대로 2019년, 2020년, 2021년의 PSR을 나타낸다. (4)~(6)열은 2019년, 2020년, 2021년에 해당하는 시가총액 계산 결과이다.

Target firm	PSR			Market Capitalization(1,000 KRW)		
	(1)	(2)	(3)	(4)	(5)	(6)
SNF	0.451	1.033	0.588	110,356,185	257,092,027	175,909,806
YNCC	0.360	0.560	0.305	1,804,275,847	2,207,028,964	1,977,101,522
LTINEOS	0.300	0.365	0.294	166,617,951	200,746,644	406,641,257
BASF	0.367	0.807	0.458	599,318,836	1,536,776,678	1,242,579,020
HTE	0.337	0.495	0.277	3,322,035,298	3,371,805,714	2,735,655,617
LTMCC	0.300	0.572	0.355	161,410,224	243,617,181	199,614,027
EIL	0.374	0.660	0.349	41,777,950	52,502,961	34,438,824
SKGEO	0.291	0.418	0.243	3,446,221,848	3,540,548,354	2,838,848,108
SKADV	0.290	0.500	0.297	226,173,721	310,751,167	266,376,441
SKPUC	0.269	0.587	0.340	291,561,550	575,848,467	253,886,508
LXMMA	0.715	0.949	1.081	476,057,684	514,407,502	813,704,171
KHMTS	0.790	1.051	1.201	546,019,812	780,988,969	1,525,830,131
PNB	0.817	1.137	1.403	1,130,381,296	1,606,141,152	3,772,097,290
SAMNAM	0.312	0.360	0.306	271,386,426	198,181,935	254,604,369
SAKM	0.639	0.856	1.002	24,914,490	34,804,716	44,468,661
SHINWN	0.358	0.409	0.370	12,703,763	14,738,208	18,105,767
DOW	0.435	0.550	0.528	202,778,910	234,526,540	253,335,267
HAEIN	0.403	0.489	0.446	21,350,306	22,511,099	37,124,982

<표 3> PSR 및 시가총액 측정(계속)

Target firm	PSR			Market Capitalization(1,000 KRW)		
	(1)	(2)	(3)	(4)	(5)	(6)
KUMJNG	0.477	0.633	0.719	9,728,682	12,440,752	19,636,854
ALMIN	0.360	0.651	0.704	25,014,124	45,703,926	58,547,446
KUDHIN	0.497	0.667	0.690	15,490,661	23,293,208	25,330,294
SINGW	0.353	0.435	0.406	6,852,699	5,966,073	5,391,345
EMTC	0.559	0.790	1.076	10,953,170	12,403,973	31,180,971
WZHIT	0.318	0.418	0.516	13,730,600	17,801,445	27,142,076
DDC	0.319	0.647	0.925	5,968,089	12,942,242	16,312,872
SNZN	0.330	0.667	0.858	11,726,986	21,621,773	40,254,280
SNKL	0.243	0.492	0.694	11,844,356	21,004,882	34,729,373
JWHIN	0.285	0.691	0.628	13,173,541	29,787,347	36,897,264
AEZIS	0.315	0.795	0.652	32,039,525	75,579,017	77,513,770
HWSNG	0.483	0.771	0.795	3,852,377	4,501,364	1,423,972
YWC	0.400	0.449	0.489	6,606,577	8,968,018	10,206,624
SILK	0.495	0.603	0.474	56,802,100	66,142,001	76,545,996
ECON	0.783	0.859	1.166	42,796,894	43,597,049	65,619,261
HWIN	0.262	0.323	0.342	5,494,225	6,525,898	8,298,708
AKEMA	0.464	0.582	0.548	42,581,771	9,581,053	10,811,858
DHPAC	0.833	0.914	1.241	22,914,903	22,832,030	35,521,523
SUNGU	0.773	0.911	0.991	17,561,803	19,576,051	27,535,481
DUBON	0.306	0.692	0.676	19,429,737	42,281,998	55,100,152
IPSI	0.756	0.825	1.100	41,687,731	44,863,924	69,416,813
NOFMT	0.912	1.001	1.359	20,692,940	19,009,466	27,797,731
NBCNA	0.844	0.926	1.257	21,232,186	22,684,062	35,613,241
SKHS	0.342	0.390	0.442	14,113,850	15,882,319	19,598,357
SAMJI	0.813	1.070	0.885	21,724,575	23,605,154	30,371,666
SAMGW	0.414	0.654	0.638	7,875,429	11,537,450	13,264,462
CMSTAR	0.838	0.934	1.210	45,323,926	47,198,987	69,418,630
ESPC	0.697	1.001	1.437	23,437,854	35,381,254	57,559,437
DHPOLI	0.717	0.804	1.018	13,841,745	15,484,360	25,006,591
DYCNE	0.440	0.681	0.479	7,465,106	10,502,211	12,310,471
FLOW	0.546	0.675	0.620	7,877,509	10,195,010	12,507,567
SILITEC	0.669	0.881	0.589	11,626,298	16,268,703	15,457,212
ACNANO	0.747	0.986	1.363	12,701,208	18,771,522	30,725,393
SAEHAN	0.668	0.748	0.938	33,464,405	38,429,180	59,209,388
LUVV	0.328	0.377	0.365	3,872,842	4,802,185	5,666,773
ESC	0.444	0.583	0.405	6,009,647	6,580,331	5,122,659
NCH	0.601	0.653	0.865	14,821,335	16,521,277	24,398,409
HAMIL	0.776	0.938	0.935	19,341,841	21,527,481	28,649,882
BNB	0.488	0.799	1.052	7,010,713	10,901,016	15,371,275
CAS	4.799	17.538	37.258	66,296,462	214,036,943	14,491,853
BSP	0.133	0.431	0.514	9,259,784	24,215,100	28,026,149

<표 3> PSR 및 시가총액 측정(계속)

Target firm	PSR			Market Capitalization(1,000 KRW)		
	(1)	(2)	(3)	(4)	(5)	(6)
KYDO	0.241	0.295	0.202	3,177,905	3,138,979	2,873,086
IPK	0.129	0.445	0.429	22,245,265	58,224,321	60,968,185
NRCOIL	0.214	0.248	0.267	21,416,674	25,282,278	32,861,410
DONGS	0.214	0.390	0.265	19,932,230	27,446,896	22,459,063
BYC	0.234	0.264	0.260	3,913,578	4,565,394	5,048,855
SDP	0.275	0.449	0.513	4,173,154	6,581,847	7,544,546
SIKA	0.191	0.330	0.308	9,537,070	14,485,941	14,332,572
JUNGS	0.247	0.304	0.352	19,361,717	23,922,982	24,709,752
NRCHEM	0.210	0.233	0.192	18,447,779	19,012,536	21,469,851
AKZO	0.209	0.256	0.316	12,305,429	14,174,420	18,165,971
DNC	0.327	0.270	0.307	14,010,717	14,080,792	17,027,242
CHUNIL	0.264	0.302	0.292	4,590,825	5,336,363	6,098,116
SAMJNG	0.148	0.335	0.238	2,744,681	5,825,695	4,092,985
DHP	0.315	0.368	0.368	11,442,252	13,873,175	12,833,300
JUNBS	0.175	0.327	0.384	8,863,167	14,919,633	22,345,066
NYC	0.242	0.322	0.335	6,201,739	8,041,600	11,195,017
HEPS	0.198	0.217	0.234	4,227,848	4,702,579	6,216,847
PNC	0.294	0.271	0.356	10,669,962	10,538,030	16,333,649
HANJIN	0.001	0.371	0.336	14,763	4,301,244	4,570,146
BGW	0.227	0.338	0.312	2,599,738	3,867,099	3,369,236
HSCHEM	0.247	0.283	0.289	5,030,175	5,875,215	7,180,639
ILSIN	0.211	0.254	0.309	6,543,154	8,142,076	11,641,888
SCT	1.167	1.806	2.223	15,112,138	25,122,244	33,763,288
OMIA	22.024	37.417	46.974	2,260,940,933	3,359,233,793	4,206,811,667
GORI	0.460	0.618	0.737	9,244,976	13,487,766	13,390,913
SVSL	2.092	3.891	1.928	139,815,308	248,775,797	158,774,981
SGCHEM	0.702	1.047	1.218	14,323,543	23,184,965	23,085,673

두 번째 시장변수인 변동성도 위와 유사한 방식으로 계산한다. 먼저, 변동성은 90일 기준으로 측정한 일일 주식 수익률의 표준편차를 연율화한 값이다. 변동성 계산 방법은 Campbell et al.(2008)에서 제시한 변동성(SIGMA) 변수를 참고하였다. 합성 타겟기업의 변동성(Volatility, 이하 Vol)은 <표 2>에서 계산한 유사기업의 포트폴리오 가중치에 실제 유사기업의 변동성을 곱한 값들의 총합으로 계산된다. 이를 수식으로 표현하면 다음과 같다.

$$Vol_{Target} = (Vol_1, Vol_2, \dots, Vol_p) \cdot \beta_{Lasso} \quad (9)$$

합성 타겟 기업의 변동성 계산 결과는 <표 4>에 표기하였다. (1)~(3)열은 차례대로 2019년, 2020년, 2021년의 변동성 계산 결과를 나타낸다. 모든 기업에서 코로나가 본격화된 2020년에 변동성이 크게 증가하는 것을 확인할 수 있다. 변동성이 100% 이상으로 높게 측정된 기업의 경우, 연구기간 동안 최소 두 번 이상 부실기업으로 분류된 적 있는 기업이다.

<표 4> 변동성 측정

<표 4>에서는 합성 타겟기업의 변동성 지표 계산 결과를 표기하였다. 변동성은 90일 기준 일일 주식 수익률의 표준편차를 연율화한 것을 의미하며, JY Campbell et al.(2008)에서 제시한 변동성(SIGMA) 변수를 참고하였다. 합성 타겟기업의 변동성(Volatility, 이하 Vol)은 <표 2>에서 측정된 타겟기업에 대한 유사기업의 포트폴리오 가중치에 실제 유사기업의 변동성을 곱한 값들의 총합으로 계산된다. (1)~(3)열은 차례대로 2019년, 2020년, 2021년의 변동성 계산 결과를 나타낸다. 모든 기업에서 코로나가 본격화된 2020년에 변동성이 크게 증가하는 것을 확인할 수 있다. 변동성이 100%이상으로 측정된 기업의 경우, 연구기간 동안 최소 두 번 이상 부실기업으로 분류된 적 있는 기업이다.

Target firm	Volatility		
	(1)	(2)	(3)
SNF	26.94	50.95	43.43
YNCC	21.77	38.78	29.30
LTINEOS	33.15	44.19	56.90
BASF	24.69	45.09	41.15
HTE	22.48	38.65	30.92
LTMCC	29.34	48.93	49.14
EIL	18.45	34.88	24.53
SKGEO	21.00	34.63	30.02
SKADV	22.89	37.91	36.02
SKPUC	21.38	38.95	35.78
LXMMA	21.93	31.39	29.74
KHMTS	23.84	34.48	32.25
PNB	21.11	29.32	33.15
SAMNAM	15.77	20.44	19.37
SAKM	18.53	26.44	25.90
SHINWN	18.54	22.19	24.37
DOW	16.36	24.94	18.71
HAEIN	17.55	24.62	20.85
KUMJNG	16.60	24.79	22.29
ALMIN	9.56	17.01	19.09
KUDHIN	18.49	32.26	21.26
SINGW	16.26	24.04	19.36
EMTC	11.46	13.99	21.63
WZHIT	10.22	11.85	16.25
DDC	20.55	46.10	63.05
SNZN	17.86	35.79	48.85
SNKL	15.29	33.87	46.32
JWHIN	12.47	23.50	41.21

<표 4> 변동성 측정(계속)

Target firm	Volatility		
	(1)	(2)	(3)
AEZIS	13.14	25.84	49.71
HWSNG	13.79	15.82	21.58
YWC	16.16	18.41	24.24
SILK	23.16	27.75	28.53
ECON	12.81	18.22	29.00
HWIN	23.86	28.53	26.74
AKEMA	16.83	15.79	26.61
DHPAC	13.63	19.38	30.85
SUNGU	13.87	17.68	27.05
DUBON	52.29	55.61	44.85
IPSI	17.36	23.69	32.45
NOFMT	14.93	21.23	33.78
NBCNA	13.81	19.64	31.26
SKHS	9.44	10.39	16.85
SAMJI	27.64	26.84	40.84
SAMGW	50.01	54.12	39.26
CMSTAR	14.00	19.42	30.65
ESPC	17.15	19.47	24.75
DHPOLI	12.82	16.90	27.31
DYCNE	20.05	20.94	22.47
FLOW	10.41	12.32	18.30
SILITEC	20.13	21.98	25.14
ACNANO	24.77	28.68	30.42
SAEHAN	13.26	16.26	27.51
LUVC	15.94	19.74	20.37
ESC	10.45	10.37	16.07
NCH	15.56	20.98	27.34
HAMIL	14.39	17.63	26.56
BNB	28.92	29.13	25.09
CAS	288.47	250.36	123.13
BSP	22.08	33.52	42.90
KYDO	15.39	19.85	26.01
IPK	23.30	30.75	43.18
NRCOIL	15.90	23.63	29.85
DONGS	23.91	24.74	36.88
BYC	15.57	21.58	25.86
SDP	27.73	40.88	53.01
SIKA	22.71	26.89	38.02
JUNGS	17.87	27.95	33.49
NRCHEM	11.56	17.27	20.82
AKZO	16.90	25.91	32.07

<표 4> 변동성 측정(계속)

Target firm	Volatility		
	(1)	(2)	(3)
DNC	19.23	28.19	33.08
CHUNIL	16.91	25.45	31.17
SAMJNG	22.06	21.53	33.43
DHP	20.23	30.92	37.30
JUNBS	21.34	30.30	40.35
NYC	20.95	28.54	36.91
HEPS	13.19	19.95	24.17
PNC	21.13	30.37	36.25
HANJIN	25.68	27.75	47.05
BGW	21.71	27.42	37.18
HSCHEM	17.49	25.92	32.66
ILSIN	17.71	26.54	33.72
SCT	38.53	45.38	50.69
OMIA	430.22	80.77	297.35
GORI	8.78	41.88	28.49
SVSL	55.46	223.49	99.91
SGCHEM	23.81	45.90	40.89

2. 랜덤 포레스트(Random Forest)를 이용한 부실예측

본 연구는 코로나-19의 영향으로 인한 재무적 부실기업을 중점적으로 분석하기 위해 다음과 같은 절차에 따라 랜덤 포레스트(Random Forest; 이하, RF) 분석을 수행하였다. 연구에는 2019년부터 2021년까지의 기업 데이터 및 CFFA로 측정된 시장변수를 활용하였다. 연구 모형의 구축을 위해 전체 데이터 중 80%를 훈련 데이터로 사용하였으며, 나머지 20%는 검증용 데이터로 사용하였다. 그후, RF 모형의 하이퍼 파라미터 튜닝을 수행하였다. 하이퍼 파라미터 튜닝은 모형의 성능과 일반화 능력을 극대화하기 위해 조정하는 작업이다. 본 연구에서는 다양한 하이퍼 파라미터 조합을 탐색하고 검증 데이터를 활용하여 모형의 성능을 평가한 결과, 다음과 같은 하이퍼 파라미터 조합을 결정하였다: n_estimators=200, max_features=0.9, max_depth=3, min_samples_leaf=4, min_samples_split=2, random_state=42.

여기서 n_estimators는 RF에 사용되는 트리의 개수를 나타낸다. max_features는 각 트리가 무작위로 선택할 수 있는 최대 특성의 비율이다. 최적의 성능을 위해 90%의 특성을 사용하도록 설정하였다. max_depth는 트리의 최대 깊이를 의미하며 이를 조절하여 모형의 복잡도를 제어하였다. min_samples_leaf 및 min_samples_split는 트리 노드의 분할을 제어하여 모형의 일반화 능력을 향상시킨다. 최소 4개의 샘플이 있는 리프 노드와 최소 2개의 샘플이 필요한

분할로 설정함으로써 모델의 안정성을 높였다. random_state는 실험을 재현 가능하게 하기 위해 랜덤 시드를 42로 설정하였다.

본 연구에서는 최종적으로 부도가 예상되는 기업을 파악하는 것이 아니라, 잠재적 부실예정인 기업을 사전에 파악하여 기업 및 채권기관의 선제적인 대응을 돕는 것을 목표로 한다. Choi et al.(2018)는 파산이나 부도와 같은 법적 절차가 시작되기 전에 채무자의 재무적 부실여부를 예측하기 위해 2년 연속으로 당기순이익이 음(-)의 값을 보이는 경우를 부실 기업으로 분류하였다. 따라서 본 연구에서는 이를 참고하여 다음 두 가지 조건 중 하나라도 만족하는 기업을 모두 잠재적 부실기업으로 정의하였다.: 1) 2년 연속 당기순이익 음(-)의 값을 갖는 경우; 2) 회계상 완전잠식에 해당하는 경우. 재정적 건전 기업과 부실 기업의 분포 비율은 <표 5>에 표기하였다.

<표 5> 부실기업 수

<표 5>은 재정적 건전 기업과 부실 기업의 비율을 나타낸다. 본 연구에서는 다음 두 조건 중 하나라도 만족하는 경우 모두 잠재적 부실기업으로 정의하였다.: 1) 2년 연속 당기순이익 음(-)의 값을 갖는 경우; 2) 회계상 완전잠식에 해당하는 경우. 잠재적 부실기업의 분포를 살펴보면 부실률은 매해 증가하여 2021년에는 부실률이 10%를 넘어서는 것을 확인할 수 있다.

Year	Non-distressed	Distressed	%
2019	80	6	6.98
2020	78	8	9.30
2021	75	11	12.79

부실예측을 위해 CFFA로 측정한 시가총액 및 변동성 지표를 비롯하여 기업의 수익성, 유동성, 활동성, 안정성, 성장성 재무 데이터 및 기타 데이터를 모두 활용하였다. 시가총액은 부실예측에 좀더 적합한 형태로서 ‘시가총액/부채’ 비율(MC/D)로 조정되었다. 이 비율은 기업이 현재 가지고 있는 부채에 대해 얼마나 잘 건딜 수 있는지를 평가하는 데 사용된다. 각 변수들에 대한 계산식은 <부록>에 표기하였다.

각 변수에 대한 기초 통계량은 <표 6>에 표기하였다. 측정값 개수(N), 평균, 표준편차, 중앙값, 최소값, 최대값을 차례로 표기하였다. 부실 예측에는 총 86개의 비상장 기업이 포함되었다. 이상치를 포함하는 변수에 대해서는 로버스트 스케일링(Robust scaling)을 진행하였다. 또한, 일부 변수들 중에서 정규분포 형태를 띄지 않고 한쪽으로 치우치는 경향이 있어 다음과 같은 작업을 진행하였다. Expense(판매관리비/매출액), CR(유동비율), NPM(당기순이익률), Quick(당좌비율), MC/D(시가총액/총부채)은 로그변환 하였으며, Intangible(무형자산비율)과 Debt(부채비율), Vol(변동성)은 박스-콕스 변환을 진행하였다. 이 과정

에서 값이 0이 나오거나 음의 왜도(Negative skew)를 보이는 경우에는 상수를 더하거나 양의 왜도(Positive skew)로 변환한 후에 변환 작업을 진행했다.

<표 6> 부실예측 변수

<표 6>은 부실예측에 사용된 각 변수의 기초 통계량을 나타낸 표이다. 부실예측을 위해 CFFA로 측정된 시가총액 및 변동성 값을 포함하여 기업의 수익성, 유동성, 활동성, 안정성, 성장성 재무 데이터 및 비재무적 데이터가 모두 활용되었다. 시가총액은 부실예측에 좀더 적합하도록 'MC/D(시가총액/부채)' 비율로 변수를 조정하였다. 이는 기업이 얼마나 많은 부채를 가지고 있으며 이에 대해 얼마나 잘 건딜 수 있는지를 평가하는 데 사용된다. 표에는 측정값 개수(N), 평균, 표준편차, 중앙값, 최소값, 최대값을 표기하였다. 부실예측에는 연구대상인 비상장 기업 86개사만 포함되었다. 이상치를 포함하는 변수에 대해서는 로버스트 스케일링(Robust scaling)을 진행하였다. 또한, 일부 변수들 중에서 정규분포 형태를 띄지 않고 한쪽으로 치우치는 경향이 있어 다음과 같은 작업을 진행하였다. Expense(판매관리비/매출액), CR(유동비율), NPM(당기순이익률), Quick(당좌비율), MC/D(시가총액/총부채)은 로그변환 하였으며, Intangible(무형자산비율)과 Debt(부채비율), Vol(변동성)은 박스-콕스 변환을 진행하였다. 이 과정에서 값이 0이 나오거나 음의 왜도(Negative skew)를 보이는 경우에는 상수를 더하거나 양의 왜도(Positive skew)로 변환한 후에 변환 작업을 진행했다.

	N	Mean	Stdev	Median	MIN	MAX
SIZE	258	0.3896	1.2383	0.00	-1.13	4.60
AGE	258	-0.0279	0.7916	0.00	-1.53	2.13
Intangible	258	0.1045	0.3721	0.00	-0.26	0.81
Expense	258	-0.0819	0.8703	0.00	-3.03	2.93
GPM	258	0.1460	0.8203	0.00	-1.72	4.22
NPM	258	1.6533	0.3619	1.66	0.00	4.20
SGR	258	0.1152	1.0527	0.00	-4.21	6.43
Debt	258	1.1556	0.1233	1.14	0.00	1.89
CR	258	0.7885	0.4380	0.69	-0.06	2.17
Cash	258	0.2054	0.7954	0.00	-0.64	4.37
ATO	258	0.1125	0.8597	0.00	-1.83	3.63
Quick	258	0.0754	0.7856	0.00	-2.11	2.10
DDP	258	24.7762	20.8292	22.90	0.00	75.81
ROIC	258	0.1275	1.3859	0.00	-9.10	8.10
MC/D	258	0.2100	0.6796	0.00	-0.80	3.05
Vol	258	0.5234	0.2215	0.53	-0.11	1.17

<표 7>은 부실예측에 사용된 변수들의 중요도를 내림차순으로 나열한 결과를 제시하고 있다. NPM(당기순이익률)이 중요도 0.5945로 가장 중요한 변수로 적용되었으며, ROIC(ROIC(총투하자본)(%))와 Cash(현금 및 현금성 자산 구성비율)가 그 뒤를 이었다. 이러한 결과는 비상장 건설화학 기업의 경우 수익성과 유동성 수준이 부실여부에 매우 중요한 영향을 미친다는 것을 시사한다. CFFA로 측정된 MC/D(시가총액/총부채)와 Vol(변동성)은 각각 5순위, 9순위의 중요도로 활용되었다.

<표 7> 변수별 중요도

<표 7>은 부실예측에 사용된 변수들의 중요도를 내림차순으로 나타낸 결과이다. 부실예측에 있어 가장 중요하게 사용된 변수는 중요도 0.5945의 NPM(당기순이익률)이며, ROIC(ROIC(총투자자본)(%))와 Cash(현금 및 현금성 자산 구성비율)가 그 뒤를 이었다. 이러한 결과는 비상장 건설화학 기업의 경우 수익성과 유동성 수준이 부실여부에 매우 중요한 영향을 미친다는 것을 시사한다. CFFA로 측정된 MC/D(시가총액/총부채)와 Vol(변동성)은 각각 5순위, 9순위의 중요도로 활용되었다.

Variable	Importance
NPM	0.5945
ROIC	0.0891
Cash	0.0888
Quick	0.0367
MC/D	0.0355
CR	0.0344
Debt	0.0328
DDP	0.0262
Vol	0.0142
ATO	0.0117
SIZE	0.0090
Expense	0.0070
GPM	0.0060
Intangible	0.0053
AGE	0.0045
SGR	0.0043

3. 성능 평가

연구 모형의 성능평가는 혼동행렬(Confusion matrix)을 기반으로 계산된 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score를 이용하였다. 특히 CFFA를 이용해 측정된 변수가 설명력 있는 변수로 사용되었는지 검증하기 위하여 MC/D와 Vol을 포함한 모델과 이를 제외한 모델로 나누어 성능을 비교 평가하였다. <표 8>은 검증용 데이터를 기반으로 한 혼동행렬 및 성능 평가 점수를 나타낸다. CFFA로 측정된 MC/D와 Vol를 포함한 모델은 모든 평가 지표에서 0.8 이상의 높은 성능 점수를 보였다. 그에 반해, MC/D와 Vol를 제외한 경우 모든 평가지표에서 성능 저하가 나타났다. 이러한 결과는 CFFA로 측정된 변수가 부실예측에서 설명력 있는 변수로 작용한다는 점을 시사한다.

특히 재현율(Recall)은 모델이 실제 부실 기업을 얼마나 잘 감지하는지를 나타내는 중요한 점수이다. 본 연구와 같이 부실 예측이나 리스크 관리를 다루는 연구에서는 이를 특히 주목할 필요가 있다. CFFA로 측정한 변수들을 포함한 모델은 재현율이 0.8로 우수한 성능을 나타내지만, 제외한 경우 0.6으로 큰 성능 저하를 보였다. 이는 CFFA로 측정한 기업의 시장 변수가 부실기업 식별 능력을 강화하는데 기여하며, 부실 관리에 유용한 변수로 활용될

수 있음을 의미한다.

<표 8> 부실예측 성능

연구 모형의 성능평가는 혼동행렬(Confusion matrix)을 기반으로 계산된 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-score를 이용하였다. 특히 CFFA를 이용해 측정된 변수가 설명력 있는 변수로 사용되었는지 검증하기 위하여 MC/D와 Vol을 포함한 모델과 이를 제외한 모델로 나누어 성능을 비교 평가하였다. <표 8>은 검증용 데이터를 기반으로 한 혼동행렬 및 성능 평가 점수를 나타낸다. CFFA로 측정된 MC/D와 Vol을 포함한 모델은 모든 평가 지표에서 0.8 이상의 높은 성능 점수를 보였다. 그에 반해, MC/D와 Vol를 제외할 경우 모든 평가지표에서 성능 저하가 나타났다. 이러한 결과는 CFFA로 측정된 변수가 부실예측에서 설명력 있는 변수로 작용한다는 점을 시사한다. 특히 재현율(Recall)은 모델이 실제 부실 기업을 얼마나 잘 감지하는지를 나타내는 중요한 점수이다. 본 연구와 같이 부실 예측이나 리스크 관리를 다루는 연구에서는 이를 특히 주목할 필요가 있다. CFFA로 측정된 변수들을 포함한 모델은 재현율이 0.8로 우수한 성능을 나타내지만, 제외한 경우 0.6으로 큰 성능 저하를 보였다. 이는 CFFA로 측정된 기업의 시장 변수가 부실기업 식별 능력을 강화하는데 기여하며, 부실 관리에 유용한 변수로 활용될 수 있음을 의미한다.

	Included	Not included
TN	46	46
FP	1	1
FN	1	2
TP	4	3
Accuracy	0.96	0.94
Precision	0.80	0.75
Recall	0.80	0.60
F1-score	0.80	0.67

V. 결 론

기존에는 제한된 데이터 가용범위로 인해 금융기관은 비상장 기업의 재무능력 평가와 자본 요건 계산에 종종 어려움이 있었다. 본 연구는 Kang(2021)의 Comparable Firm-Finding Algorithm(CFFA)를 활용하여 비상장 기업의 가치를 평가하고 부실예측에 필요한 정보를 보완하는 방법을 제시하였다. CFFA는 기존의 유사기업 분석법(Comparable Company Analysis)을 일반화한 가치평가 체계로서, 회귀분석을 기반으로 하여 보다 폭넓은 적용이 가능하도록 설계되었다.

본 연구에서 CFFA를 활용하여 도출된 유사기업의 포트폴리오 가중치는 대부분 0.6 이상의 높은 R^2 값을 보였다. 이를 기반으로 측정된 ‘시가총액’과 ‘변동성’ 지표는 이후 진행된 부실예측에서 각각 5순위(시가총액/총부채 비율), 9순위(변동성)의 중요도로 활용되었다. 랜덤 포레스트 기반의 부실예측을 진행한 결과, 검증용 데이터의 성능 평가에서 CFFA로 측정된 변수를 모델에 포함한 경우 모든 평가 지표에서 0.8 이상의 높은 성능을 보였다. 이에 비해 CFFA로 측정된 변수를 제외할 경우 모든 평가 지표에서 성능 저하가 발생한다는

것을 확인하였다. 이는 CFFA를 통해 측정된 변수가 부실예측에 유의미한 변수로 작용하였음을 시사한다. 특히 성능의 차이는 재현율(Recall)에서 가장 크게 드러났다. 이를 통해 CFFA로 측정한 시장 변수를 포함할 경우 재무비율만 사용했을 때보다 부실 예측의 분류 능력을 더 향상시킬 수 있다는 결론을 얻을 수 있었다.

이번 연구의 주요 기여는 다음과 같다. CFFA를 통한 기업 정보의 보강으로 금융기관이 부실 및 리스크 관리 능력을 향상시키는 데 도움을 줄 수 있다는 점이다. 또한, 기업의 관점에서도 CFFA를 활용하여 내부 자본 건전성을 보다 적극적으로 관리할 수 있는 방안을 제시하였다는 점이다. 다만, 이 연구는 국내 건설화학 제조업에 대한 한정된 범위의 연구로 이루어져 있기에, 추후 연구에서는 더 다양한 기업을 대상으로 연구 대상 범위를 확장하고자 한다.

참 고 문 헌

- 김구희, 원유호, 이주형, “건설산업의 특성을 고려한 워크아웃기업의 가치평가”, 한국산학기술학회논문지, 제15권 제8호, 2014, 4863-4873.
- 김성진, 안현철, “기업신용등급 예측을 위한 랜덤 포레스트의 응용”, 산업혁신연구, 제32권 제1호, 2016, 187-211.
- 김주은, “주가매출비율(PSR)의 가치지표로서의 검증”, 기업경영연구, 제29권 제3호, 2022, 101-124.
- 권누리, 김영민, 최광신, “거시경제 변수를 고려한 한국기업부도 모형 구축 방법 연구”, 한국데이터정보과학회지, 제30권 제5호, 2019, 1037-1050.
- 류두진, 박대현, “산업별 특성과 기업 간 상호작용을 고려한 4 차 산업기술기업의 가치평가”, 금융공학연구, 제20권 제2호, 2021, 1-19.
- 송혁준, 오웅락, “학술연구 : 벤처기업의 비상장주식 평가방법에 관한 연구”, 기업가정신과 벤처연구(JSBI)(구 벤처경영연구), 제10권 제3호, 2007, 85-103.
- 이정환, 조진형, “머신러닝과 비재무적 정보를 이용한 부실 확률 예측 모형 : 탄소배출정보, ESG 성과, 애널리스트 정보를 중심으로”, 한국금융연구원, 연구보고서, 2023, 1-64.
- 이재득, “릿지 회귀와 라쏘 회귀 모형에 의한 부산 전략산업의 지역경제 효과에 대한 머신러닝 예측”, 한국향만경제학회지, 제37권 제1호, 2021, 197-215.
- 이지혜, “건설업 수익성 기여요인 분석 및 활용 방안”, 한국건설산업연구원, 연구보고서, 2020, 1-125.
- 전성일, 이병원, “무형자산성 지출의 기업가치관련성에 관한 연구”, 중소기업연구, 제25권 제2호, 2003, 327-352.
- Alford, A. W., “The Effect of the Set of Comparable Firms on the Accuracy of the Price-Earnings Valuation Method,” *Journal of Accounting Research*, 30(1), (1992), 94-108.
- Altman, E. I., “Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy,” *The Journal of Finance*, 23(4), (1968), 589-609.
- Barboza, F., H. Kimura, and E. Altman, “Machine Learning Models and Bankruptcy Prediction,” *Expert Systems with Applications*, 83, (2017), 405-417.
- Bee, T. K., “Valuations of Chemical Companies and Distributors: Comparable Metrics and Precedent Transactions,” *The Journal of Private Equity*, 21(4), (2018), 87-100.

- Bowman, R. G. and S. R. Bush, "Using Comparable Companies to Estimate the Betas of Private Companies," *Journal of Applied Finance, Forthcoming*, Available at SSRN: <https://papers.ssrn.com/abstract=956443>, 2007.
- Breiman, L., "Random Forests," *Machine Learning*, 45(1), (2001), 5-32.
- Campbell, J. Y., J. Hilscher, and J. Szilagyi, "In Search of Distress Risk," *The Journal of Finance*, 63(6), (2008), 2899-2939.
- Choi, H., H. Son, and C. Kim, "Predicting Financial Distress of Contractors in the Construction Industry Using Ensemble Learning," *Expert Systems with Applications*, 110, (2018), 1-10.
- Dichev, I. D., "Is the Risk of Bankruptcy a Systematic Risk?," *The Journal of Finance*, 53(3), (1998), 1131-1147.
- Hernandez Tinoco, M. and N. Wilson, "Financial Distress and Bankruptcy Prediction among Listed Companies Using Accounting, Market and Macroeconomic Variables," *International Review of Financial Analysis*, 30, (2013), 394-419.
- How, J., J. Lam, and J. Yeo, "The Use of the Comparable Firm Approach in Valuing Australian IPOs," *International Review of Financial Analysis*, 16(2), (2007), 99-115.
- Kang, H. G., "Strategically Comparable Firm-Finding Algorithm (CFFA): A Strategic Management Approach to Valuation," Available at SSRN: <https://ssrn.com/abstract=3959510> or <http://dx.doi.org/10.2139/ssrn.3959510>, 2021.
- Malakauskas, A. and A. Lakštutienė, "Financial Distress Prediction for Small and Medium Enterprises Using Machine Learning Techniques," *Engineering Economics*, 32(1), (2021), 4-14.
- Minton, B. A., C. M. Schrand, and B. R. Walther, "The Role of Volatility in Forecasting," *Review of Accounting Studies*, 7(2), (2002), 195-215.
- Ohlson, J. A., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," *Journal of Accounting Research*, 18(1), (1980), 109-131.
- Shumway, T., "Forecasting Bankruptcy More Accurately: A Simple Hazard Model," *The Journal of Business*, 74(1), (2001), 101-124.
- Steiger, F., "The Validity of Company Valuation Using Discounted Cash Flow methods," arXiv preprint arXiv:1003.4881, 2010.
- Tibshirani, R., "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), (1996), 267-288.

<부 록>

A. 변동성 계산식

구 분	계산식	참고논문
변동성	$SIGMA_{i,t-1,t-3} = \left(252 \times \frac{1}{N-1} \sum_{k \in \{t-1,t-2,t-3\}} r_{i,k}^2 \right)^{\frac{1}{2}}$	Campbell et al.(2008)

B. 부실예측 변수 계산식

구 분		계산식
수익성	GPM	매출총이익(매출액-매출원가) / 매출액*100
	NPM	당기순이익 / 매출액*100
	ROIC	세후 영업이익 / 총투하자본*100
유동성	CR	유동자산 / 유동부채*100
	Quick	당좌자산 / 유동부채*100
	Cash	(현금+현금성자산) / 유동부채*100
활동성	ATO	매출액 / 총자산
	Expense	판매관리비 / 매출액*100
안정성	Debt	부채 / 자본*100
	DDP	차입금 및 사채총액 / 총 부채
성장성	SGR	(매출액 _t - 매출액 _{t-1}) / 매출액 _{t-1} *100
	Intangible	무형자산 / 총자산*100
비재무	AGE	
	SIZE	ln(총자산)
시장가치	MC/D	
	Vol	

THE KOREAN JOURNAL OF FINANCIAL MANAGEMENT
Volume 40, Number 6, December 2023

Financial Distress Prediction of Unlisted Construction Chemical Firms Using CFFA and Random Forest

Seoyeon Jo* · MyeongSu Choi** · Hyoung-Goo Kang***

〈Abstract〉

This study uses the Comparable Firm Finding Algorithm (CFFA) and Random Forest to predict financial distress in non-listed construction chemical firms. Using CFFA, the study complements the limited information on non-listed firms by measuring their market variables, which have been difficult to obtain. In particular, to analyze the impact of the COVID-19 shock, corporate data from 2019 to 2021 was used to predict financial distress. As a result of the performance evaluation, when variables measured by CFFA were included, high scores of over 0.8 was shown in all evaluation indicators. In comparison, when variables measured by CFFA were excluded, performance deteriorated in all indicators. The difference in performance was greatest in recall. Therefore, it was confirmed that the variables measured by CFFA can be applied as explanatory variables in predicting insolvency and strengthen the ability to identify distressed companies. Based on these findings, this study is expected to contribute to enabling firms and creditor institutions to proactively respond to insolvency management.

Keywords : CFFA, Valuation, Financial distress prediction, Machine learning, Random Forest

* First Author, School of Business, Hanyang University, E-mail: chosoj@hanyang.ac.kr

** Co-Author, Adjunct Professor, School of Business, Hanyang University, E-mail: kidsjjang@hanyang.ac.kr

*** Corresponding author, Associate Professor, School of Business, Hanyang University,
E-mail: hyoungkang@hanyang.ac.kr