KLI

정책자료
2023-02

자연어 기반 특허와 노동시장 성과 분석

방형준 · 곽도원 · 이가현

한국노동연구원

목 차

표 목 차

그림목차

요 약

본 보고서는 특허의 경제적 효과와 노동시장 성과를 전통적인 계량경제학적 방법과 자연어 분석 두 가지를 사용하여 분석을 시도하였다. 특허는 연구개발 활동의 결과물 중 하나로, 측정 불가능한 연구개발 활동의 성과를 가늠케 하는 주요한 지표 중 하나이다. 따라서 경제학에서는 특허를 연구개발 활동이나 생산성 등에 대한 간접적인 지표로 사용하곤 한다.

본 연구에서는 연구개발 활동의 성과나 투자 정도를 특허로 판별하지 않고, 한국평가데이터에서 작성한 "R&D 활동성 지수"와 기업의 연구개발 투자 비용으로 측정하였다. 분석 결과에 따르면, 연구개발 활동은 연구개발 투자 비용으로 측정하였을 때, 1인당 매출액으로 측정한 생산성과 양의 상관관계를 가지는 반면, R&D 활동성 지수는 음의 관계를 가지는 것으로 나타났다. 이러한 상반된 결과는 첫째로 두 가지 지표가 높은 상관관계를 보임에도 불구하고 측정 방법이 다르다는 차이에서 기인할 수 있다. 또한 R&D 활동성 지수가 연구개발 활동의 장기 지속성을 반영하는 반면, 연구개발 투자 비용은 즉각적인 지출이기에 여기서 기인하는 차이도 있을 수 있다. 기업의 경영성과나 생산성이 높아짐에 따라 연구개발에 대한 지출을 늘리는 것이 마치 연구개발 투자 비용과 생산성 간의 정의 상관관계로 측정되었을 수 있기 때문이다.

하지만 종사자 수로 측정한 고용지표의 경우 R&D 활동성 지수와 연구개발 투자 비용 모두와 양의 상관관계를 보이는 것으로 분석되었다. 이는 연구개발 전문 인력이 기존의 사무직이나 생산직 인력과는 다른 학력 및 숙련 요건을 필요로 하므로 신규채용을 늘려야만 연구개발 활동을 전개할 수 있어서일 수 있다. 또 한편으로는 규모가 커지고 그에 따라 종사자 수가 늘어난 기업에서 사업의 확장 및 시장 점유율 유지 등을 위

해 연구개발 활동을 전개할 필요를 느끼는 역의 인과관계가 존재할 수도 있다. 본 연구에서는 두 가지 경로 중 무엇이 더 큰 요인으로 작용하는지에 대해서는 자료의 한계로 밝히지 못하였다.

한편 R&D 활동성 지수와 연구개발 투자 비용 두 지표 모두에서 일정한 수준 이하의 점수나 투자 비용이 기업의 생산성이나 매출액, 고용지표 등과 무관한 관계를 가지는 것을 확인하였다. 이는 연구개발 활동이 실질적인 성과를 내기 위해서는 일정한 규모 이상으로 이루어져야 함을 시사한다.

제4장에서는 우리나라의 대표적인 연구개발 지원 정책인 국가연구개발 지원사업을 유사한 각도에서 평가하여 보았다. 그 결과 국가연구개발 지원사업에 참여한 기업에서는 특허 출원이나 등록의 성과에서 증가 효과를 관찰하였으나, 사업 참여 이전부터 존재한 선행적인 증가 추세까지 고려한다면 실질적인 특허 등록 수의 증가 효과는 찾을 수 없었다. 이러한 결과는 특허 등록이 연구개발 활동으로부터 긴 시간을 두고 이루어지는 시차 효과에서 기인했을 수도 있고, 연구개발 활동에 관심이 많은 기업이 사업에 적극적으로 지원하고 참여하는 표본 선택으로 인한 것일 수도 있다.

국가연구개발 지원사업의 참여가 당기순이익이나 매출 같은 경영 지표와 음의 상관관계를 가지는 것으로 나타났는데, 이는 연구개발 활동이 투자 초기에는 비용의 성격을 가지기 때문으로 보인다. 반면 국가연구개발 지원사업은 고용 측면에서 사업 참여 기업에서 참여 이후 뚜렷한 양의 효과를 내는 것을 확인할 수 있었다. 이는 국가연구개발사업 참여에 따른 큰 규모의 연구개발 활동 지원으로 인해 기업이 장기적인 생산성 증대를 기대해서 사업 규모를 키우고 종사자 수를 늘렸기 때문일 수 있다.

한편 특허는 초록과 본문 등의 각종 정보가 자연어로 이루어져 있는바, 이를 바탕으로 특허의 신규성, 피인용 횟수로 측정한 특허의 가치, 고용 변화와 연계된 특허의 특징을 자연어로 분석하였다. 결과에 따르면, 한국어로 된 특허의 신규성을 한국어 기반 자연어 모형으로 측정한

결과는 영어 위주로 학습된 다언어 모형에 비해 월등히 좋은 성과를 거두었다. 이는 한국어가 다른 언어와 비교하여 뚜렷히 구별되는 자연어적 특성을 가지고 있기 때문으로 보이며, 따라서 특허 심사나 특허 분석에 있어 한국어 기반 자연어 모형을 구축할 필요성을 제시한다 하겠다.

특허의 가치는 명시적으로 드러나는 지표가 아니기 때문에 정확한 특허의 가치를 산정하거나 측정하는 것은 어려운 작업이다. 이로 인해 많은 연구에서는 특허의 가치를 피인용 횟수로 측정한다. 즉, 인용 횟수가 많을수록 특허가 더 높은 가치를 가진다는 것이다. 본 연구에서는 실제 피인용 횟수가 적절한 생산성의 측정지표가 될 수 있는지를 간접적으로 파악하기 위해서 동일한 법인에서 과거 20년간 출원한 특허를 피인용 횟수별로 구분하여 자연어의 특성을 살펴보았다. 그 결과 업종에 따른 자연어상의 차이는 뚜렷하게 드러났으나, 피인용 횟수에 따른 차이를 관찰하기는 쉽지 않았다. 다만, 1,000회 이상 인용된 특허의 경우 어느 정도 자연어상의 차이를 보였기 때문에, 향후 피인용 횟수가 아주 많은 특허와 그렇지 않은 특허를 비교하는 후속 연구에서 보다 정확한 결론을 얻을 수 있을 것으로 보인다.

마지막으로 특허로 인한 고용효과를 자연어상의 특징으로 식별할 수 있는지 알아보기 위해서, 고용 증가나 감소가 큰 기계・장비 산업 법인에 대해서 고용이 적은 시기와 많은 시기를 놓고 특허의 차이를 살펴보았다. 그 결과 시기별 차이는 뚜렷하게 관찰되었으나, 고용이 늘어나거나 줄어든 데 대한 특허의 자연어 차이는 뚜렷하게 나타나지 않았다.

이러한 연구 결과들은 다음의 정책적 시사점을 제공한다.

첫째로, 연구개발 활동이 실제 생산성 증가 등의 성과를 내기 위해서는 일정한 규모 이상의 투자를 필요로 한다는 점이다. 이러한 정책적 시사점이 대기업 위주의 연구개발 지원 활동을 전개해야 함을 의미하지만은 않는다. 중소기업에 대한 대규모 연구개발 활동 지원을 통해 기업의 규모를 키운다면 이후에는 자체적인 연구개발 활동이 가능하므로 지속적으로 새로운 중소기업의 스케일업을 유도할 수 있다. 다만, 이를 위해서는 다양한 연구개발 관련 사업을 진행하기보다는 소수의 집중적인 대

규모 사업 집행이 보다 효율적일 수 있음을 뜻한다.

또한 고용과 연구개발 활동 간의 선순환 구조가 발견됨에 따라, 양질의 일자리 창출을 위해서 연구개발 활동을 장려할 필요가 있다. 일반적인 생산직이나 사무직 종사자들은 연구개발 전문 인력으로 전환되기 어려우므로, 장기에 걸쳐 지속적으로 연구개발 활동을 전개하려는 기업은 필연적으로 연구개발 전문 인력을 채용할 수밖에 없다. 일반적으로 연구개발 직군은 요구되는 학력 수준이 높고 임금과 근로 여건이 괜찮은 양질의 일자리이다. 따라서 노동시장에서 양질의 일자리를 많이 공급하기 위한 정책의 하나로 연구개발 활동을 지원하는 정책을 펼칠 수 있다.

마지막으로 한국어 기반 자연어 분석 모형은 한국어로 된 문서를 분석함에 있어 다언어 모형에 비해 뚜렷한 장점을 가짐이 확인되었다. 앞으로는 특허 심사나 R&D의 수행, 특허와 연구개발 활동에 대한 평가 작업 등에서 자연어 분석의 도움을 받을 필요가 있을 수 있으므로, 심사관과 기업의 편의를 도모하고 국내의 자연어 기반 연구를 장려하기 위해서라도 한국어 기반 자연어 모형을 구축하는 것이 필요할 것이다.

제 1 장
서 론

제1절 연구 배경

기업이 경영 활동을 영위할 때 고려하는 주요 목표 중 하나는 이윤 창출이다. 수학적으로 이윤은 기업의 사업 매출액에서 인건비 및 임대료를 포함한 사업에 소요되는 모든 비용을 제외하고 남은 금액으로 정의할 수 있다. 이윤은 시장에서 경쟁하는 다른 경제 주체들과 비교하여 기업이 얼마나 저렴하게 좋은 제품을 생산해 내는가가 가장 큰 영향을 미친다. 이때 저렴하게 좋은 제품을 생산할 수 있는 능력을 경제학적으로 그리고 수학적으로 측정한 지표가 생산성이다. 따라서 생산성이 높은 기업일수록 일반적으로 높은 이윤을 창출하는데, 문제는 생산성에는 이윤이나 매출, 혹은 고용원 수 등 수학적으로 정확하게 측정하거나 모두가 보편적으로 수용하는 계측 방법이 없다는 것이다. 따라서 경제학자들은 기업의 생산성을 측정하기 위한 다양한 변수를 사용하는데, 그중 하나가 바로 특허이다.

기업이 특허를 취득하거나 사용하기 위해서는 연구개발(R&D) 활동을 벌여 특허를 출원할 수도 있고, 혹은 기(旣) 출원된 특허를 구매하거나 혹은 특허 사용에 대한 대가를 치르기도 한다. 이 중에서 특허를 출원하는 행위는 특허를 구매하거나 비용을 지불하고 사는 것과 다른 성격을 띠고 있다. 이미 개발된 특허를 구매하거나 혹은 사용료를 지불하는 것은 특허의 존재 여

부나 가치에 있어서 불확실성이 적다. 이미 출원된 특허이므로 특허의 내용과 성격, 활용법 등에 대해서 알고 있는 상황에서 특허의 가격이나 비용에 대해서만 협상하면 될 뿐이며, 쌍방이 특허의 내용이나 가치에 대해 이미 알고 있는 상황이므로 비용이나 가격에 있어서 이견(異見)이 크지 않을 가능성이 높다. 그러나 기업이 특허를 개발한다면 큰 불확실성을 감내하는 것이다.

특허를 개발하기 위해서는 필연적으로 연구개발을 수행해야 한다. 그러나 연구개발 활동이 반드시 일정 수준 이상의 가치를 가진, 혹은 투입된 개발 비용 이상의 가치를 가지는 특허를 출원하는 결과로 이어진다고 담보할 수 없다. 다행히 성공적인 연구개발이 이루어진다면 투입된 비용보다 더 높은 수익을 거둘 수 있겠으나, 일부 특허는 지출된 비용을 고려하였을 때 현저히 낮은 가치를 창출하기도 한다. 심지어 어떠한 경우에는 연구개발 활동이 실패로 끝나 아무런 특허나 성과물을 얻지 못하고 종료되기도 하며, 다른 경쟁 기업들이 먼저 연구 성과를 생산하여 그간의 연구개발 활동을 종료하고 다른 방향으로 연구를 시행해야만 하는 상황에 직면하기도 한다.

연구개발 활동 혹은 특허 출원은 이러한 불확실성이 큰 행위이지만, 그렇다고 연구개발 활동을 전혀 하지 않는다면 기업은 비싼 가격을 지불하고 특허를 사용하거나 혹은 특허에 대한 접근이나 사용을 거부당하여 아예 제품을 생산할 수 없는 상황에 이르기도 한다. 또한 특허는 일단 출원되면 그 이후는 특허를 유지하기 위한 수십만 원의 비용만 감당하면 되는 반면, 특허를 활용하거나 다른 기업에 사용하게 할 때 추가적으로 소요되는 비용이 없다. 따라서 양질의 특허를 다수 보유하면 기업의 수익성이 개선되며, 기업의 생산성 역시 상승하는 것으로 나타난다. 이러한 연구개발 활동의 성과는 거시경제 전반에 양의 외부효과를 창출하기 때문에 정부는 연구개발 활동을 장려하기 위해 다양한 노력을 기울이게 된다.

경제학에서 인위적인 외부의 개입이 없는 경우 일반적으로 양의 외부효과가 있는 경제 활동은 시장의 최적 수준보다 적게 공급되고 음의 외부효과가 있는 경제 활동은 과다하게 공급된다고 알려져 있다. 따라서 연구개발 활동 역시도 정부의 인위적인 장려 활동이 없다면 시장 전체 혹은 국민경제를 고려하였을 때 과소공급될 가능성이 있다. 따라서 정부는 국민경제 전체에서 최적인 수준보다 적게 이루어질 연구개발 활동을 장려하기 위해 다양한

정책과 제도를 전개하고 있다.

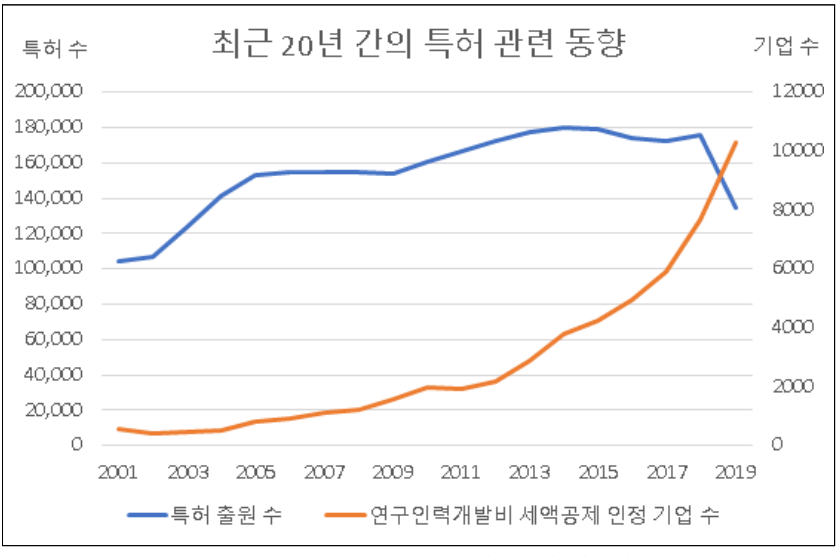
이러한 대표적인 사례로 국가연구개발사업과 연구소 전담부서 사업을 꼽을 수 있다. 국가연구개발사업은 '국가연구개발사업 관리 등에 관한 규정'에 따라 중앙정부에서 국가적 수요나 필요가 제기된다고 판단되는 연구개발 과제에 대해서 해당 연구개발 활동에 소요되는 비용의 전액 또는 일부를 출연하거나 공공기금을 활용하여 지원하는 사업이다. 연구소 전담부서 사업은 과학기술정보통신부에서 과학기술 분야 또는 서비스 분야의 연구개발 활동을 촉진하기 위해 일정한 요건을 갖춘 민간 사업체의 연구소 혹은 연구전담 부서를 설치한 기업에 대해서 취득세와 법인세 등의 소득세를 75% 감면하고, 기술보증기금의 보증한도를 증액하는 한편, 미취업 청년을 고용하였을 때 해당 종사자의 인건비에 대해서 최대 1년간 50%를 지원받을 수 있다. 더하여 창업 이후 3년 이내에 벤처기업 인증까지 받게 된다면 5년간 법인세의 반을 경감받는다.

이러한 직간접적인 특허 출원 활동에 대한 장려 혹은 연구개발 활동에 대한 지원의 성과 등으로 인해 우리나라의 특허 출원 수는 2000년대에 꾸준히 증가하는 추세를 보여왔다. [그림 1-1]은 2001년 이후 각 연도별로 우리나라에서 출원된 특허의 수와 관련 정책 중 하나인 연구인력 개발비 세액공제 인정 기업 수를 표시한 것이다. 전반적으로 우상향하는 추세, 즉 특허 출원 수가 증가하는 추세에 있음이 확인되고 있다.

그러나 모든 특허가 동일한 가치를 가지지는 않는다. 어떠한 특허는 사용료가 더 비싸거나 고가에 소유권이 거래되거나, 최종 생산품의 생산에 지대한 기여를 하기도 하는 반면, 일부 특허는 전혀 활용되지 않거나 인용되지 않는다. 일반 상품의 경우 판매 가격이라는 명확한 가치 산정 기준이 존재하는 반면, 특허의 경우에는 이러한 객관적인 기준이 존재하지 않는다. 따라서 특허의 가치를 어떻게 산정할 것인가 하는 문제는 경제학자들이 안고 있는 오랜 숙제 중 하나이다.

이에 대해 경제학자들이 고심한 결과는 몇 가지가 있다. 하나는 특허의 매매 가격이나 사용료를 기준으로 하는 것이다. 이 기준은 시장에서 거래되는 재화나 용역처럼 특허의 가격이 시장에서 명시적으로 확인된 경우이다. 더군다나 거래가 실제로 성사되었기 때문에 수요와 공급 양측에서 모두 해

[그림 1-1] 연도별 우리나라 특허 출원 건수 및 연구인력개발 세액공제 인정 기업 수



자료 : 한국특허정보원 자료 및 연구소 전담부서 신고관리 시스템 자료를 바탕으로 저자 가공.

당 특허의 가치에 대해서 비슷한 견해를 가지고 있다는 의미이기도 하다. 그러나 이 기준을 실제 특허의 가치 산정에 적용하는 데에는 치명적인 문제점이 존재하는데, 바로 거래되거나 혹은 사용료를 지불하고 사용되는 특허 개수가 실제 출원되는 특허의 수와 비교하여 매우 적다는 점이다. 따라서 이 방법을 사용하였을 때 다수의 특허가 가치가 없거나 매우 적다는 결론에 다다르게 되는데, 이것이 합리적이라 할 수는 없다.

다른 방법은 연도별 특허의 갱신료를 가지고 계산하는 것이다. 특허 보유주는 특허의 내용과 가치에 대해서 다른 경제주체들보다 더 많은 정보를 가지고 있을 것이며, 따라서 특허권 존속 기간 동안의 기대 수익 혹은 예상 수입에 대한 정보도 가지고 있을 것이다. 만일 어느 시점에서 현재 가치로 환산한 특허의 잔존 가치가 갱신료보다 더 낮다면 해당 경제주체는 특허의 갱신을 포기할 것이다. 이를 바탕으로 특허의 가치를 계산해 볼 수 있다. 그러나 이러한 계산법은 갱신료가 특허 가치에 따라서 다르게 부과되지 않는 상황에서 분석에 어려움을 준다. 물론 갱신료에 차이가 없다 하더라도 특허의 잔존 가치에 따라서 어떠한 특허는 보다 일찍 갱신되지 않고 어떠한 특허는

오래토록 갱신될 텐데, 출원 시기가 오래되지 않은 특허의 경우에는 정확한 가치를 산정하기가 어렵다. 우리나라의 경우, 특허 유지 비용에 대해서 지원이 이루어지기도 하는데, 이것이 특허의 가치 등과는 무관하게 사업체 규모에 따라서 이루어지는 경우가 많다 보니 이러한 계산을 더욱 어렵게 한다.

따라서 본 연구에서는 특허의 가치와 기업의 연구개발 활동에 대해서 기존과는 다른 측정 방법을 시도하고자 한다. 우선 연구개발 활동을 두 가지 측면에서 분석할 것이다. 하나는 국가의 연구개발 활동 지원 정책에 참여했는지 여부이다. 대표적으로 국가연구개발사업이 있다. 이 사업은 특정 R&D 주제에 대해 연구개발을 진행하는 사업체에 해당 연구개발에 소요되는 비용의 일부 혹은 전액을 직접 국가가 지원하는데, 실패했을 경우 이후의 참여에 제한이 가해질 수 있으므로 연구개발을 담당한 사업체는 해당 연구개발을 성공시키기 위해 노력할 유인이 존재한다. 따라서 사업체는 연구개발을 성공시키기 위해 많은 노력을 기울이고 이로 인해서 연구개발 활동이 활발해질 수 있다. 다른 하나는 한국평가데이터에서 개발한 "R&D 활동성 지수"를 사용하는 것이다. 연구개발 활동은 매출이나 영업이익과 같이 객관적으로 측정할 수 있는 단위가 존재하지 않는다는 문제가 있으며, 더하여 연구개발 활동의 성과물은 실제 해당 활동이 이루어지는 시점과 시차가 존재한다는 약점이 있다. 이를 보완하기 위해서 "R&D 활동성 지수"는 사업체의 각종 재무정보를 이용하여 성공 여부와 무관하게 실제 기업이 전개하는 R&D 활동의 정도가 어느 수준인지를 측정하려는 것이다. 이를 바탕으로 실제 기업의 연구개발 활동이 기업의 생산성이나 고용지표와 어떠한 관계를 가지는지를 살펴보고자 한다.

특허의 가치에 대해서는 기존의 연구들이 주로 인용 횟수를 가지고 살펴보았는데, 인용 횟수가 특허의 가치와 직접 연결된다 할 수는 없다. 그러나 현실적으로 특허의 가치를 직접적으로 측정할 방법이 없다면, 객관적인 계량 지표로서 특허의 인용 횟수 이외의 다른 대안을 찾기도 쉽지 않다. 따라서 특허의 초록을 대상으로 하여 인용 횟수가 높은 특허와 그렇지 않은 특허에 대해서 자연어상으로 어떠한 차이를 발견할 수 있는지를 알아보고자 한다. 즉, 특허의 가치가 자연어상으로도 드러나는지 알 수 있다면, 출원 초기 특허의 경우 다양한 지표를 활용하여도 짧은 관측 시점으로 인하여 정확

한 특허 가치를 파악하거나 예상하기가 어려우나, 자연어 분석은 그에 대해서 한 가지 방향을 제시할 수 있을 것이다.

제2절 연구의 구성

본 보고서는 다음과 같은 네 개의 장으로 구성된다. 우선 연구의 목적과 필요성을 소개하고 전체 구성을 제시하는 서론이다.

제2장에서는 선행 연구를 소개하고, 아울러 실증 분석에서 사용할 자료에 대해서 설명하고자 한다. 이어서 제3장에서는 기업의 연구개발 활동이 실제 기업의 생산성을 높이는지, 높인다면 얼마나 높이는지, 그리고 기업의 연구개발 활동은 어떠한 고용효과를 창출하는지 분석하였다. 제4장에서는 연구개발 활동을 장려하기 위한 정부정책의 효과를 실증적으로 평가하기 위해 국가연구개발사업에 참여했는지 여부에 따른 기업 성과를 분석하였다. 평가지표로는 출원한 특허 수, 당기순이익, 매출, 납입 자본금 등의 경영성과와 고용원 수로 측정한 노동시장 성과 변수를 살펴보았다. 제3장과 제4장은 본 보고서가 살펴보고자 하는 연구개발 활동과 고용 간의 관계를 보다 정확하게 추정할 수 있게 해줄 것이다.

제5장에서는 자연어 분석에 초점을 맞추어서 개별 특허들의 초록을 이용하여 자연어 분석을 실시하고자 한다. 초점은 자연어 분석을 이용하여 특허의 신규성을 얼마나 정확하게 예측할 수 있는지를 탐색함으로써, 이후 특허의 신규성을 판별하며 기업의 생산성을 더 잘 측정할 수 있는 방법을 탐색할 뿐만 아니라, 연구개발 활동의 수준이나 고용지표의 차이가 특허의 성격에도 차이를 가져오는지를 자연어 분석을 통해 살펴보고자 한다. 이를 바탕으로 자연어 분석이 특허의 고용효과 분석을 위한 측정지표로서 기능할 수 있는지의 가능성까지 살펴보고자 한다.

마지막으로 제6장에서는 본 보고서의 연구 내용을 정리하고 결과에 대한 개괄적인 해석을 제공할 것이다. 아울러 이를 바탕으로 연구개발 활동을 장려하기 위한 정책 방향과 특허 제도에 대한 정책적 시사점도 모색해 볼 것이다.

제 2 장
선행 연구 및 자료

제1절 서 론

본 장은 기업의 연구개발 활동이 생산성이나 고용에 미친 효과에 대한 선행 연구를 탐색하고, 아울러 다음 두 장에서 설명할 생산성과 고용효과에 대한 실증 분석에서 사용한 자료를 소개하고자 한다. 경제 성장을 위한 핵심 요인으로는 자본과 같은 생산 요소의 축적과 혁신적인 생산 방식 도입에 의한 기술 요인을 들 수 있다. 특히, 지속적인 자본 축적을 통해 경제 성장을 달성한 고소득 국가의 경우 일정 수준 이상의 성장이 이루어진 후에는 자본의 한계생산 체감에 의해 요소 축적을 통한 성장원(成長原)을 대부분 소진한다. 따라서 고소득 국가의 경제가 지속적으로 성장하기 위해서는 자본과 같은 생산 요소의 축적 이외에도 기술의 역할이 중요해지며 그 중요성은 경제가 성장할수록 커진다.

경제학 이론에서는 새로운 아이디어, 발명, 혁신, 생산 방식 개선 등을 통해 동일한 생산 요소를 투입하더라도 더 많은 최종 생산물의 산출이 가능하면 이를 기술 진보라 부른다. 생산 요소를 더 효율적으로 조합할 수 있는 새로운 아이디어, 발명, 혁신이 활발하게 이루어지는 곳에서는 기술 진보를 통한 경제 성장의 달성이 가능하다.

경제학에서 기술의 생성을 위한 투입은 연구 및 개발에 대한 투자의 형태

로 나타나며, 새로운 기술이라는 산출물은 다양한 지표를 사용하여 수량화 할 수 있다. 다만, 다양한 기술개발 지표 중에서 가장 널리 사용되는 것은 연구 및 개발에 대한 투자이다. 우리나라 정부는 과거부터 기술입국 등의 표어를 내걸며 기술개발의 중요성을 강조하여 연구개발 활동에 대한 투자를 지속적으로 확대하여 왔으며, 민간 분야에서는 기업도 자체적으로 대규모의 연구개발 인력의 확보 및 투자를 지속적으로 시행하고 있다. 정부에서는 다양한 수단을 활용하여 기술 혁신을 촉진하려 하는데 가장 대표적인 형태의 정부 개입으로 연구개발 (R&D) 활동을 위한 보조금 지급 정책을 거론할 수 있다.

연구개발 투자의 효율성에 대해서는 다양한 연구가 이루어져 왔는데, 연구개발 투자의 효율성에 대한 연구에 따르면 산업 분야에 따라서 그리고 다른 경제적 환경 요인에 의해 그 효과가 상이하게 나타날 수 있다. 예를 들어, 연구개발에 대한 투자로 인해 기술 혁신이 이루어지면 이는 노동시장에도 영향을 미쳐 노동시장의 대표적 성과지표인 고용량과 임금, 궁극적으로는 임금 분포에도 영향을 미칠 수 있다. 연구개발 투자가 경제 전반에 미치는 영향에 대해 방대한 선행 연구가 존재하는데, 본 장에서는 이러한 선행 연구들을 정리하여 소개하고자 한다.

이어서 3절에서는 본격적인 분석에 앞서 본 보고서에서 실증 분석을 수행하기 위해 활용한 다양한 데이터를 소개하고 해당 데이터를 어떻게 가공하여 사용했는지를 설명하고자 한다.

제2절 선행 연구

연구개발 활동이 기업이나 국가의 생산성 혹은 고용에 미치는 영향을 탐색한 대표적인 국내 연구로는 김진영(2012)를 꼽을 수 있다. 해당 연구는 한국의 특허 데이터로부터 구축된 발명자 개인 단위의 패널자료를 이용하여 각 발명자의 기술 생산성을 결정하는 요인을 탐색하기 위한 요인 분석을 실시하였다. 발명자 개인의 생산성은 특허 출원 수를 대리지표로 사용하였고

대표적인 생산성 결정 요인으로는 발명자의 연령 등에 중점을 두어 관계를 추정하였다. 추정 방법론은 특허 출원 수가 0을 포함하는 양의 자연수 값을 갖는 성질을 고려하여 음이항(Negative binomial) 방법의 통계 분석을 실시하였다. 결과에 따르면 특허 생산성은 발명가의 나이와 역U자형 관계를 보이며, 개발 생산성의 정점은 대략 30~34세에 도달하는 것으로 밝혀졌다. 성별에 따른 이질적 효과 분석에서는 특허 출원 수를 활용하였는데, 기술개발에 있어 여성의 생산성은 남성의 생산성보다 평균적으로 낮게 나타났으나, 특허의 질을 통제하고 나면 성별 차이는 사라졌다. 연도별 생산성을 보면 특허 출원 수는 지속적으로 증가하였고, 젊은 세대의 특허 출원 생산성이 나이든 세대보다 높게 나타났다. 특허 수와 특허 청구 수 간에는 역의 상관관계가 있으며, 규모에 따른 생산성에서는 규모가 더 큰 조직에서 더 많은 특허를 생산하는 것으로 관찰되었다.

조상섭 외(2003)는 OECD 25개국 특허 자료를 이용하여 내국인과 외국인을 구분해 특허 출원의 노동생산성에 대한 기여도를 추정함으로써 자국에서 생산된 지식과 다른 나라에서 생산된 지식의 파급 효과를 분석하였다. 연구 결과를 요약하면 다음과 같다. 첫째, 노동생산성과 내국인 특허 출원량 및 외국인 특허 출원량 변수에 대해서 IPS 패널 단위근 검정 결과 단위근 존재에 대한 명확한 증거를 발견하지 못했다. 둘째, 패널 OLS 추정량을 사용하여 노동생산성에 대한 기여도를 분석한 결과에 따르면, 내국인 특허 출원 양의 노동생산성 기여도는 0.18이고 외국인 특허 수의 노동생산성 기여도는 0.15로 나타났다. 이는 외국인에 의해 출원된 특허 수가 내국인의 특허 수에 근접하여 외국인에 의한 특허 개발도 노동생산성의 증가에 상당한 기여를 한다는 사실을 보여준다. 마지막으로 노동생산성에 대한 전체적인 지식 파급 효과(0.331)는 직접생산 투자 요소인 물적 자본량의 평균 기여도 0.3 (Smolny, 2000)에 비하여 작지 않게 나타났다. 또한 지식의 파급 효과가 경제의 지속적인 성장을 위하여 중요하다는 점을 보여주었으며, 지속적인 경제 성장을 위한 기술 진보에 있어서 내국의 연구개발 결과물뿐만 아니라 지식 선진국의 개발 성과를 적극적으로 받아들여 지식의 파급 효과를 크게 하는 것이 중요함을 보여주었다. 따라서 외국인에 의한 지식의 파급을 극대화하기 위해 외국인 특허 유치 정책을 펼치는 것이 도움이 될 수 있음을 보여

주었다.

서환주와 이영수(2005)는 국가 간 자료를 사용하여 기술 혁신과 생산성 사이에 중요한 관련이 있음을 확인했다. 논문에서는 특허 건수를 지식의 대리변수로 사용하여 지식과 기술의 변화가 국가 간 성장 격차를 설명하는 데 중요한 역할을 한다는 것을 보여주었다. 또한 고정자본에 대한 투자가 국가 간 성장 격차 확대에서 핵심적인 역할을 한다는 것을 확인했으며, 고정자본에 대한 투자와 경제 성장 사이에 누적되는 인과관계가 있다는 것을 보여주었다. 이러한 누적적인 인과관계로 인해 국가 간 경제 성장의 격차는 시간이 지남에 따라 확대될 수 있는 것으로 나타났다. 특허권 강화의 경제적 효과를 기술 혁신과 고정자본에 대한 투자의 효과로 구분하여 연구한 결과, 기술 혁신의 경제 성장 촉진 효과는 통계적으로 유의미하지 않았지만 고정자본에 대한 투자는 경제 성장과 양의 상관관계를 가지는 것으로 나타났다. 따라서 특허권 강화가 고정자본에 대한 투자를 촉진하기는 하지만, 기술혁신으로 이어지는 결과를 보여주지는 못했다. 실증분석의 추정 결과에 따르면, 국가 간 기술개발과 경제 성장 경로가 다양하며, 고정자본 투자, 기술개발 및 사회자본은 서로 상호작용을 하며 연결되어 있다. 그러므로 각 국가는 지적재산권 강화를 통해 기술개발에 친화적인 경제적 환경을 조성하고 고정자본에 대한 투자를 촉진해야 하며, 진학률 향상을 통해 인적자본의 질을 높이고, 아울러 소득 불평등 완화 정책을 통해 사회적 결속력을 제고해야만 지속적인 경제 성장으로 이어질 수 있는 기술개발이 가능한 것으로 주장하였다.

김정언 외(2007)는 한국 데이터를 활용하여 특허권 강화 정책이 기술혁신과 경제성장에 미치는 영향을 분석하였다. 지식 기반 경제에서는 지식 창출과 기술 확산이 지속적인 경제 성장을 위해 중요하므로, 지적재산을 보호를 위해 지적재산권 관련 법률을 강화할 필요를 강조했고, 이를 위해 최근 많은 선진국에서 특허법을 개혁하고 있음을 보여주었다. 한국에서도 특허권 강화 정책이 도입되고 이를 통해 특허에 대한 보호의 수준이 높아졌으며, 궁극적으로 특허 출원 수의 증가로 이어졌음을 보였다. 주요 실증분석 결과로는, 특허 출원에 미치는 효과가 기업 특성 변수와 비교했을 때 그 중요성에 있어 특허권을 강화하는 정책이 통계적으로 유의하게 나타나지 않았다.

기업의 내부 특성 및 시장의 경쟁 정도, 기술의 파급 효과 등 다른 결정 요인을 고려할 때 특허권을 강화하는 것이 기술 혁신이나 기술 진보를 촉진하는 효과는 통계적으로 유의하지 않았다. 그러나 산업별 분석 결과에 따르면, IT 산업에서의 기술 혁신이 다른 산업에 긍정적인 파급 효과를 내는 것으로 나타났다. 즉, IT 산업의 기술 혁신은 국내 기업들의 노동생산성을 향상시키는 긍정적인 영향을 가져왔다. 연구 결과에 기반하여, 특허권 강화가 지적재산 보호뿐만 아니라 다른 보조적인 환경과 조합될 때에만 최대의 효과를 발휘할 수 있다고 추론하였다.

오동현 · 김소영(2015)은 우리나라 중소기업에서 특허권이 생산성, 기술 혁신 및 기술 추격에 미치는 영향을 정량적으로 분석하였다. 2001년부터 2012년까지 미국 특허 상표국에 등록된 특허와 한국 기업데이터의 재무 자료를 병합하여 자료를 구성하였으며, 확률적 변경 분석법을 이용하여 우리나라 중소기업의 생산성 변화율에 특허권이 미치는 영향을 분석하였다. 분석 결과에 따르면 생산성과 규모 효과, 그리고 분배 효율성은 특허권을 보유하고 있는 기업에서 더 높게 나타났으나 기술 혁신에 있어서는 특허권을 보유하고 있는 중소기업에서 더 나은 결과가 나타나지는 않았다. 이러한 결과를 바탕으로 우수한 기술 수준을 보유한 기업 역시 지속적인 기술 혁신을 위한 노력이 필요함을 역설하였다.

장선미(2020)는 기술 축적을 위한 대표적인 투입 요소인 연구개발 투자와 그 산출물인 특허 출원 수를 이용하여 한국의 제조업체를 대상으로 산업별 기술 수준을 측정하고, 이를 산업 생산성과 함께 분석하여 산업 수준에서 기술의 영향력을 살펴보았다. 이 연구는 기술 축적이 경제 성장에 미치는 영향을 분석하는 것을 목적으로 하는데, 이때 기술 축적의 대리지표가 특허 출원 수이며 경제 성장의 대리지표로 산업 생산성을 채택하여 양자 간의 관계를 분석했다. 결과에 따르면 연구개발에 대한 투자는 현재의 투자와 누적 투자 모두 지식 산출에 유의미한 양의 영향을 미치는 것으로 나타났다. 연구개발 투자 유량과 연구개발 투자 누적량을 함께 고려하면 투자 유량만 유의하고 누적량의 유의성은 발견되지 않았다. 그리고 지식의 투입 요소인 연구개발 투자의 증가는 생산성에 긍정적인 영향을 미치지 못하지만, 연구개발 활동의 결과물인 특허의 경우 생산성 증가에는 유의한 양의 영향을 미치는

것으로 나타났다. 즉, 산업 내에서 생산성이 증가하기 위해서는 지식의 축적이 필요한데 연구개발 자체보다는 특허와 같은 실질적인 산출물이 생산성의 증가를 불러오는 요인이었다.

장지연 외(2022)는 특허 자료를 기술 변화의 대리지표로 활용하여 특허와 직업을 연계해 직업별 특허량 변수를 생성한 후 이를 이용하여 기술 변화에 따른 임금 및 고용 변화를 분석하였다. 이를 위해 기계학습 기법인 FastText를 사용하여 O*NET의 텍스트와 CPC 텍스트 간의 유사성을 계산하고, 이를 토대로 세분류(4digit) 기준의 직업별 특허량 변수를 생성했다. 생성한 특허량 변수를 이용해서 우리나라 노동시장 변수에 매칭하기 위해 미국표준직업분류-국제표준직업분류-한국표준직업분류 간의 연계표에 기반하여 세분류 기준의 직업별 특허량을 소분류 기준의 직업별 특허량으로 다시 집계했다. 그리고 생성된 소분류 기준의 직업별 특허량을 하나의 변수로 하여 기술 변화가 임금 및 고용 변화에 미치는 효과를 회귀분석하였다. 분석 결과 특허량으로 대리된 기술 노출도와 임금 수준과의 관계는 성비, 교육 수준, 연령 등의 인구·사회적 변수의 값에 따라 동일한 직업에서도 성(性), 임금 규모, 숙련 수준, 연령대별로 상이하게 나타났다. 보다 구체적으로 살펴보면, 첫째로 전체 특허 출원량과 신기술 관련 특허 출원량은 2015년 이후에 감소하는 추세를 보였고, 이러한 추세의 변동은 조선과 자동차부품 산업의 구조조정과 2020년의 코로나19로부터 영향을 받은 것으로 해석하였다. 둘째, 특허량과 노동시장 변수 간에 비선형관계가 존재하는데, 특허량의 노동시장 변수에의 효과는 임금, 성별, 교육 수준, 연령 등 개인의 특성에 따라 달랐다. 셋째, 고정효과를 이용한 패널모형 분석에 따르면, 다른 선행 연구들의 결과와 비슷하게, 특허량이 임금과 고용에 오히려 부정적인 영향을 미치는 것으로 나타났다. 이러한 결과들에 기반하여 해당 연구는 다음과 같은 개선점을 제시하고 있다. 첫째, 직업별 특허량을 생성하는 과정에서 직업 내 업무 특성을 충분히 고려할 필요가 있는데, 이것이 변수 생성에 잘 반영되지 못했다고 밝혔다. 둘째, 본 연구는 기술 전반과 새로운 기술 분야를 다루었지만, 후자에서는 특허량이 부족하여 주로 전자 분야에 초점이 맞춰져 분석이 이루어졌다는 한계가 있는데, 신기술 분야에 집중적으로 정책 자금을 지원하는 최근의 경향을 고려할 때 다양한 기술 분야를 포괄하는 분석이 필요

하다. 셋째, 기술 변화의 노동시장 영향은 복잡하며, 불확실성이 존재하지만, 실증 분석에서는 불확실성의 요인을 구체적으로 수량화하여 분석하지 않아 연구 결과를 해석하는 데 신중을 기할 필요가 있음을 언급하였다.

정성철 외(2004)는 특허가 부가가치 생산에 영향을 주는 경로를 두 가지로 가정하고 각 경로의 효과를 분석하였다. 첫째 경로는 특허가 기업의 기술혁신 활동을 촉진하여 결과적으로 부가가치 생산의 중요한 요소인 지식의 저량을 늘려줌으로써 부가가치 생산에 기여하는 것이다. 둘째 경로는 특허가 기술 정보의 확산을 촉진함으로써 지식 저량의 활용을 확산하여 부가가치 생산에 기여할 수 있다. 분석 결과에 따르면 특허 제도가 연구개발 투자를 촉진함으로써 지식 저량의 증가에 기여하고, 우리나라 제조업의 경우 특허 제도가 시장 경쟁 촉진에도 기여하는 것을 보여주었다.

이경곤(2018)은 임금 불평등과 기술 사용 증가 사이의 관계를 살펴보았다. 이전 연구들은 높은 수준의 기술 변화가 진행된 산업에서 더 높은 임금을 보고하며, 이러한 임금 불평등의 주요 원인으로 기술 편향적 기술 진보를 지적하였는데, 이와 관련하여 특허 출원 데이터를 활용하여 194개 산업에서의 기술 변화를 분석함으로써, 기술 변화의 수준과 임금구조 간의 관계를 새로운 시각으로 제시하였다. 실증 분석 결과, 기술 변화에 의해 고학력 노동자의 소득은 증가하는 반면, 고령의 숙련 노동자에게는 손실이 발생하며, 이로써 임금 불평등과 기술 변화 간의 밀접한 관련을 확인하였다. 이러한 연구 결과는 4차 산업혁명과 같은 신기술 발전이 미래에 소득 양극화를 더욱 심화시킬 가능성을 시사하며, 이에 대응하기 위한 인적자본 형성과 교육정책의 중요성을 강조하였다.

양성준 · 김동현(2021)은 최근 4차 산업혁명으로 떠오른 인공지능(AI)의 발전이 노동시장에 미칠 영향에 대해 다루고 있다. AI에 취약한 직종에 대해서 직업적 특성을 중심으로 조사하고 해당 직종들의 지리적 분포를 살펴보았다. 구체적으로 우리나라의 특허 데이터를 분석하여 국내 AI 기술의 특성을 파악하고, 이와 관련된 직무와 직업을 식별하였다. 식별 결과 AI에 취약한 직종은 주로 반복적 업무를 수행하는 직종으로 나타났고, 반면 물리적 업무나 창의성을 요구하는 업무를 하는 직종은 AI의 영향에서 비교적 자유로운 것으로 확인되었다. 또한 AI의 영향은 지역마다 상이한 양상으로 나타났

는데, 이를 살펴보기 위해 서울과 진주를 비교하였다. 예를 들면, 서울에서는 노동자의 업무 형태에 따라 양극화 현상이 두드러졌으나, 진주에서는 양극화 현상이 나타나지 않았다. 지역별로 AI의 영향이 다르게 나타나므로, 이에 기반하여 정부정책이 지역을 중심으로 노동시장의 구조를 고려해 전개될 필요성을 제기하였다. 또한 AI의 기술적 특성을 고려한 새로운 형태의 업무와 노동환경을 형성해야 양극화 현상을 완화할 수 있음도 역설하였다.

특허나 연구개발 활동에 관한 해외 연구로 우선 Dechezleprêtre 외(2019)를 거론할 수 있다. 해당 논문은 고임금이 자동화를 촉진하는지에 대해 연구하면서, 특정 키워드의 빈도수를 활용하여 기계를 이용한 자동화의 정도를 식별하는 새로운 자동화 측정지표를 개발하였다. 미국의 업종 간 분석에서 자동화 측정지표 값이 커질 때 일상적인 작업이 감소하는 것을 상관관계 분석을 통해 보여주었다. 그리고 자동화와 관련된 특허에 관한 기업 패널 데이터를 구축하고, 41개 국가의 거시경제 데이터와 지리적 특허 이력 정보를 결합하였다. 구축한 데이터를 활용한 실증분석 결과 저임금 노동자의 증가가 자동화 혁신과 양의 상관관계가 있음을 보여주었다. 반면 고임금 노동자의 증가는 자동화 혁신과 음의 상관관계가 나타남을 보여주었다.

Gambardella(2005)는 기업이 연구개발 활동을 전개할 유인이 공급사슬 내에서 기업의 위치에 따라 달라질 수 있음을 보였다. 기존의 선행연구에 의하면, 많은 기술 전문 기업들은, 주로 대규모 제조회사들인 기술 구매자들과 비교하면 기업 규모는 작지만 구매자들에게 기술을 공급하는 경우에는 높은 협상력은 갖는다. 강력한 지식재산권이 부재(不在)한 경우, 기술 공급자는 자신들의 기술을 최종 제품에 내재화하여 그 기술에 대한 임대료를 확보하기 위해 하류(생산이 분절화되어 여러 단계로 구성된 상황에서 하부 생산 단계)로 통합하려고 시도할 수 있다. 이는 Gans and Stern(2003)이 언급한 포인트이다. 이러한 논거에 기반하여 해당 논문에 따르면 약한 지적재산권은 분업을 제한하며 기술 전문 소규모 기업 간의 기업 병합이 나타날 수 있고, 이로 인해 기술 전문가들이 자체적으로 통합 기업이 되거나 혹은 적은 수의 독립적인 공급 업체로 뭉치기 때문에 기술 수요자이자 구매자인 제조 기업들이 상류로 통합해야 하는 상황을 야기한다고 주장하였다.

Kaiser(2008)은 노동자 이동이 지식과 기술의 이전 수단으로서 중요하게

나타남을 실증분석을 통해 보여주었다. 유럽 특허청과 연계된 덴마크 기업에 대한 고용주-근로자 등록 데이터에서 덴마크의 노동력을 "R&D 노동자"와 "비 R&D 노동자"로 분류한 후, "비 R&D 노동자"에서 "R&D 노동자"로 이동한 노동자들이 계속해서 "비 R&D 노동자"에 남아 있는 노동자들보다 특허 출원 활동에 더 많이 기여하는 것을 보여주었다. 또한, 특허를 출원한 기업에서 이전에 고용되었던 R&D 노동자인 "특허 노출 노동자"는 이러한 경험이 없는 R&D 노동자보다 특허 출원 활동에 더 적극적이라는 것도 보여주었다. 특허 노출 R&D 합류자는 특허 출원의 관점에서 가장 생산적인 집단이었으며, 1999년 이전에 특허를 출원한 기업에서 이러한 유형의 노동자를 추가로 고용하는 고용주는 특허 출원 수가 증가한다는 것도 관찰되었다. 평균적으로 연간 특허 출원 수 증가분의 14%가 특허 노출 노동자로부터 비롯된 것으로 분석되었다. 또한, R&D 노동자의 이동이 이들을 수용한 기업과 기존에 이들이 재직하던 기업 간의 공동 특허 활동을 증가시켜, 덴마크 경제 전체의 신기술 개발에도 긍정적인 영향을 발휘한다는 것을 보여주었다.

Danzer 외(2020)는 노동과 자본 간의 대체 가능성을 제안하는 경제이론과 노동공급량의 변화가 노동을 절약하는 기술의 증가에 미치는 효과가 실증분석을 통해 검증되는지 분석하였다. 이 논문은 1990년대와 2000년대 독일의 이민자 배치 정책을 활용하여 지역 노동공급의 외생적 변화가 자동화에 미치는 영향을 분석하였다. 이중차분법을 활용하여 고숙련 노동자와 저숙련 노동자 1,000명을 기준으로 여기서 추가로 고용된 노동자 1명당 자동화를 위한 특허가 0.05개 감소하는 것과 상관관계가 있음을 보여주었다. 이러한 특허 개발에서의 음의 효과는 신규 노동자가 유입된 지 2년이 경과한 후에 가장 현저하게 나타났다. 다만, 이러한 결과는 저숙련 노동자가 많은 산업에 국한되어 나타났다. 해당 논문은 독일에서의 이민 정책을 활용하여 노동공급이 자동화에 영향을 미치는 인과관계를 준실험 방법에 기반하여 연구하였고, 이를 통해 이민 정책을 이용한 노동과 기술개발 혁신 간의 관계를 연구함에 있어 새로운 접근법을 제공하였다.

노동시장과 기업의 연구개발 활동을 연결 지은 연구로는 Autor 외(2023)도 있다. 이 논문은 기업 규모별로 이질적인 효과를 고려하여 규모가 큰 미국 기업(슈퍼스타)의 혁신 활동이 전체 8개년 동안 노동시장에 미칠 가능성

을 고려하였다. 슈퍼스타 기업은 혁신의 상당 부분을 창출하며, 다른 기업의 혁신에 비해 차별적이고 더 큰 영향을 미치는 것으로 나타났다. 새로운 특허 수준에 대한 측정치를 활용하여, 슈퍼스타 기업의 기술 혁신은 특히 최근 몇 십 년 동안 다른 기업이 창출한 혁신에 비해 노동 보완적인 경향을 가지는 것으로 나타났다. 그러나 이러한 기술과 노동 간의 보완적인 관계는 기술의 숙련도에 따라 상이하게 나타났다. 상위기업에서 나타나는 고용과 임금 간의 양의 관계는 고임금 직업군의 노동자에게 한정되었다. 이는 슈퍼스타 기업의 기술 혁신이 고숙련 노동자와 저숙련 노동자가 노동시장에서 상반된 효과를 낼 수 있음을 보여준다.

Alesina 외(2018)에서는 노동 절약적인 기술에서 노동시장 정책의 다양한 영향을 분석하였다. 노동시장에서의 규제는 기술 발전에 따른 고숙련 노동의 저숙련 노동 대비 숙련 프리미엄을 감소시켰다. 따라서 노동을 절약하고 자본의 의존도를 높이는 기술은 노동규제가 더 엄격한 국가에서 저숙련 노동자에게 더 큰 제약을 부과할 것으로 예상되었다. 따라서, 노동규제의 정도가 높은 곳에서는 고숙련 분야의 기술 수준을 뒤처지게 하지만, 저숙련 분야에서는 오히려 기술적으로 더 발전하게 만드는 것으로 나타났다.

이러한 방대한 선행 연구에도 불구하고 기술혁신과 노동시장 성과물 간의 관계에 대한 연구는 여전히 더 필요하다. 첫째, 기존의 선행 연구는 기술혁신과 노동시장 성과물 간의 관계에 있어 일관된 결과를 보여주지 않고 있다. 이는 기술혁신과 노동시장의 성과물 간에 다양한 측면이 존재함을 의미하며, 이에 따라 기술혁신이 노동시장 성과에 미치는 영향을 측정할 때 이질적 효과를 고려할 필요가 있음을 알 수 있다. 또한, 현 시점에서 정부는 대규모의 기술혁신을 위한 국가지원 사업을 특정 산업에 선택하여 집중하고 있으므로 이러한 산업에서도 기술혁신과 노동시장 성과물의 관계가 기존 문헌의 연구 결과와 동일하게 유지되는지 최근의 데이터를 통해 검증할 필요가 있다. 특히, 인공지능이나 이차전지, 5G 및 6G와 같이 고숙련 노동자가 집중되어 있는 산업에서도 특허와 노동시장 성과 변수 간에 기존의 관계가 동일하게 도출되는지 아니면 새로운 결과가 나타나는지에 대한 엄밀하고 실증적인 검증이 필요하다.

제3절 자 료

1. 한국기업데이터

본 장에서 분석에 활용한 주요 자료는 한국평가데이터에서 발행하는 “한국기업데이터”이다. 한국기업데이터는 기업의 업종, 설립일자, 기업의 소재지, 상시 근로자 수, 사업자등록번호를 정보 제공 일자와 함께 제공하여, 각 회계연도별로 기업의 각종 재무 정보 및 고용 현황을 담고 있다. 이를 바탕으로 패널자료를 구축하기 위해 한국기업데이터의 사업체 고유번호와 자료의 제공 연도를 이용하였으며, 보조적으로 다른 지표와의 연결을 위해 사업자등록번호 역시 사용하였다.

한국기업데이터는 각 기업의 연도별 주요 재무 정보를 담고 있는데, 여기에는 기업의 총자산, 자본총계, 매출액, 영업이익, 당기순이익을 포함하고 있어, 기업의 경영성과나 생산성을 계산하는 데 활용하였다. ‘주요 재무 정보’에서 제공하는 연구개발비 역시 기업의 연구개발 활동에 대한 정보로 활용하였다.

한국기업데이터는 기업의 고용현황에 대해 ‘인원현황’이라는 별도의 표에서 정보를 제공한다. 해당 표는 성별 상시 근로자 수, 성별 총급여 지급액 및 평균 급여액, 성별 평균 근속연수를 담고 있으나, 연구개발비와 마찬가지로 약 반 정도의 기업에 대해서 정보가 없거나 0으로 기재되어 정확성이 떨어진다고 판단하였다. 따라서 기업의 고용현황에 대해서는 ‘기업 개요’ 표에서 제공하는 상시 근로자 수만을 활용하였다. 이로 인해서 성별 근로자 수의 비율이나 평균 임금, 기업의 인건비 등에 대한 정보를 활용할 수 없게 됨은 아쉬우나, 전체 데이터의 반 이상을 활용치 못하게 되는 점, 활용 가능한 데이터의 대표성, 가중치 계산 등의 문제가 발생하기에 부득이 이러한 선택을 하였다.

한편, 한 회계연도에 하나의 기업에 대해서 두 개 이상의 한국기업데이터 정보가 기재된 경우가 있다. 이러한 중복 기재의 비율은 높지는 않으나, 주

로 재무제표를 수정했거나 추가적인 공시를 한 경우에 해당하므로 중복 기재의 경우에는 나중 일자의 정보를 정확한 것으로 간주하여 자료를 정리해 활용했다.

2. 연구개발 관련 자료

가. 국가연구개발 지원 사업 참여 기업 명단

한국평가데이터에서는 각 연도별로 국가연구개발 지원사업 참여 기업 명단을 데이터베이스화해 놓았다. 하지만 전체 국가연구개발사업 참여 기업 명단을 획득하기 위해서 국가과학기술지식정보를 활용하여 국가연구개발 지원사업을 수행한 기업과 사업수행 기간을 식별하였다. 사업수행에 대한 자료는 연/월/일 단위로 아주 상세하게 제공되나 다른 자료와의 병합을 위해 연 단위 자료로 변환하였다. 예를 들면, 한 해를 기준으로 기업별로 국가연구개발 지원사업의 지원을 받았는지 여부와 지원을 받은 연구사업의 총 숫자를 수량화하였다.

다만 국가연구개발 지원사업 참여 기업 정보의 경우 2010년 이후의 자료에서만 결측치 없이 충분히 관측되었다. 따라서 한국기업데이터는 그 이전 자료부터 존재하지만 연구개발 관련 정책에 대한 자료와의 연계성을 위해 정책 평가 부분에서는 2010년 이후의 자료만 사용하여 분석 기간을 2010~2021년으로 제한하였다.

나. 특허 출원 자료

본 연구에서 활용한 특허 출원 자료는 한국기업데이터의 24번째 항목인 지적재산권에서 세 가지 형태의 지적재산권(특허, 실용안장, 디자인)에 대해 출원 번호를 보유한 기업(출원, 공개, 등록 모두를 포함)을 특허 출원한 기업으로 식별해 내었다. 예를 들면 한 해 동안 자료에 나타나는 출원번호 건수를 통해 특허 출원 수를 계산하였다.

특허 출원 자료 역시 변수는 기업의 일 단위 및 월 단위의 미시자료를 연

단위로 집계하여 패널데이터를 구축해 한국기업데이터와 정합하는 방향으로 가공하였다.

다. R&D 활동성 지수

기업의 연구개발 활동은, 앞서 언급했다시피, 실패할 수도 있고 혹은 성공했다 하더라도 그 성과가 특허 출원과 같이 명시적으로 관찰 가능한 형태로 나타나지 않을 수도 있다. 또한 연구개발 활동으로 동일한 수의 특허를 출원했다 하더라도 특허의 가치는 모두 동일하지 않아서 어떠한 기업은 보다 생산적인 연구개발 활동을 수행한 반면, 어떠한 기업의 연구개발 활동은 가치가 낮은 특허만을 양산했을 가능성도 있다. 이처럼 기업의 연구개발 활동은 직접 관찰 가능하지 않은 데다가 그 정도를 측정하는 것 역시 난제(難題)이다. 따라서 기업의 연구개발 활동 정도를 측정하기 위해서는 새로운 지표가 필요한데, 본 장에서는 'R&D 활동성 지수'를 사용하고자 한다.

한국평가데이터에 따르면 R&D 활동성 지수는 "기업의 R&D 환경, R&D에 대한 재정적/인적 투자, 산출물을 통해 추정한 기업의 R&D 활동이 활발한 정도"로 정의된다. R&D 활동성 지수는 기업이 꾸준히 연구개발 활동을 전개하고 투자를 단행하고 있는지 장기적인 관점에서 확인하기 위한 변수이다. R&D 활동성 지수는 '연구개발 환경', '연구개발 투입', '연구개발 성과' 세 가지 지표로 구성되며, 한국평가데이터가 수집하고 확보한 기업의 연도별 연구개발비, 특허의 등록 추이, 국가 R&D 과제 수행 이력, R&D 인력 비율 등의 정보를 활용하여 산출된다.

R&D 활동성 지수를 산출하기 위해 활용하는 자료의 유형은 다음과 같다.

1) 기업 보유 특허

R&D 활동성 지수의 산출에서 가장 중요한 역할을 하는 요인은 기업이 보유하고 있는 특허이다. 특허에 대해서는 출원 특허의 수와 등록 특허의 수를 모두 고려하여 특허청의 공식적 인증을 통해 최종 산출물로 확정된 연구개발 활동뿐만 아니라 최종 산출물로 인정받기 위한 과정에 있는 특허의 수까지 고려함으로써 지수가 단순한 결과물의 지수화에 그치는 것을 막으려

하고 있다. 또한 특허의 출원이 매년 지속적으로 이루어지고 있는지도 살펴봄으로써 연구개발 활동이 장기적으로 계속되는지를 간접적으로나마 확인하고 있다.

지수 산출에 있어서 특허는 단순히 양적 지표만 반영하지 않고 질적 측면도 고려하려 하고 있다. 특허의 질적 가치를 측정하기 위해 기업이 보유하고 있는 개별 특허에 대해서 시장에서 산정된 가치가 있다면 해당 가치를 반영하고, 특허와 관련된 권리 및 한국평가데이터 내부의 TCB(Technology Credit Bureau)에서 산정한 특허의 가치 및 기술 점수까지 반영하여 현재 기업의 기술 수준을 측정하였다. 또한 가치가 있는 특허일수록 출원 후 계속해서 갱신료를 지불하여 권리를 유지하려 할 것이므로 등록된 특허 중 권리 해제 이전에 계속해서 권리를 유지하고 있는 특허의 비율이 어느 정도 되는가도 계산에서 주요 요인으로 반영된다. 또한 기업이 양적 지표에만 치중하여 제대로 된 연구개발 없이 특허의 수를 늘려감으로써 R&D 활동성 지수가 높게 나타나는 것을 방지하기 위해 출원한 특허 중 실제 특허청에 의해 등록되는, 다시 말해서 출원한 특허 중 신규성을 인정받은 특허의 수가 차지하는 비율도 지수 계산에서 참고하고 있다.

2) 연구개발 비용

기업이 연구개발 활동을 활발히 한다면 당연히 그에 수반되는 연구개발 비용 역시 발생할 수밖에 없다. 유사한 맥락에서 연구개발 활동이 지속적으로 이루어진다면 연구개발 비용 역시 지속적으로 지출되는 것이 필연적이다. 따라서 R&D 활동성 지수 산출에는 기업의 재무재표나 경영 활동에서 연구개발 비용이 지불되고 있는지 여부, 그리고 기업의 매출이나 전체 비용에서 연구개발비가 차지하는 비율은 어느 정도 되는지가 지수 산출에서 고려된다. R&D 활동성 지수에서는 일회성의 단발적인 대규모 연구개발 활동보다는 규모가 작더라도 지속되는 연구개발 활동에 더 큰 가치를 두고 있다. 따라서 연구개발 비용이 차지하는 비율도 중요하지만 이러한 연구개발 관련 비용의 투입이 지속적으로 이루어지고 있는지, 즉 연구개발 비용 지출의 연속성 여부도 반영하고 있다.

3) 연구개발 인력

기업의 연구개발 활동은 재원(財源)뿐만 아니라 인력의 투입도 필요로 한다. 기업이 연구개발 활동을 활발하게 전개한다면 기업의 총고용에서 연구개발 인력이 일정한 수준의 비중을 차지할 것이다. 그러나 연구개발을 전혀 수행하지 않거나 혹은 단발적으로 수행하는 기업의 경우에는 연구개발 인력이 없거나 혹은 지속적인 연구개발 인력의 고용을 관찰할 수 없을 것이다. 따라서 R&D 활동성 지수 산출 시, 연구개발 수행 여부와 연속성, 그리고 그 정도를 비용 측면만이 아니라 인력 측면에서도 관찰하고 있다.

4) 국가연구개발사업 참여 여부

기업이 실제로 연구개발 활동에 적극적이라면 국가연구개발사업에 참여하거나 했을 가능성이 높다. 국가연구개발사업은 중앙 행정 부서가 필요하다고 판단한 연구개발 활동을 민간기업이 수행하면 해당 연구개발 비용의 전액 혹은 일부를 정부에서 부담한다. 기업이 연구개발 활동에 관심이 없다면 우연히 수행하는 사업이 국가연구개발사업의 주제와 일치할 가능성도 낮을 뿐더러, 설령 일치한다 하여도 연구개발 활동 이력의 부실로 인해 국가연구개발사업의 수행 기관으로 지정되기도 쉽지 않다. 연구개발 활동을 등한시하는 기업은 어떠한 연구개발 주제가 국가연구개발사업에 선정되었는지도 잘 알지 못할 것이다. 또한 국가연구개발사업의 지원금은 최종 산출물이 애초에 설정한 기준을 통과했는지, 그리고 실제 해당 사업이 목표를 달성했는지에 따라서 지불되거나 최악의 경우 환수되는 경우도 있으므로 해당 주제의 연구개발 활동을 수행할 의사가 크지 않다면 참여할 유인도 적다. 그러므로 국가연구개발사업에 참여한 기업들은 일반적으로 연구개발 활동을 활발히 하며, 실제 해당 연구개발을 성공시킬 의지가 강하다고 할 수 있다.

따라서 R&D 활동성 지수 산출에서 국가연구개발사업은 몇 가지 측정지표로 반영된다. 국가연구개발사업은 사업 공고문을 통해 해당 사업의 상세한 내용과 목표, 그리고 연구비에 대해서 자세히 알 수 있다. 이를 바탕으로 해당 연구개발 활동의 연구 내용이나 설정된 연구비를 통해 중요도나 가치를 간접적으로 추산하여 지수 산출 과정에서 고려한다.

5) 기업 특성

R&D 활동성 지수는 연구개발 환경도 고려하는바, 연구개발 활동과 직접적으로 연결되는 항목들뿐만 아니라 그 외의 항목들 중에서도 연구개발에 우호적인 방향으로 작용할 수 있는 다양한 요인을 최대한 고려하여 반영하였다. 예를 들어, 기업이 연구소나 연구 기관을 보유하고 있거나 운영하고 있는지 여부는 연구개발 부서의 규모나 기업의 관심 정도를 보여주는 지표로 사용할 수 있다. 또한 이노비즈나 메인비즈, 혹은 벤처기업 인증 등 다양한 기술 인증을 가지고 있다면, 그러한 활동의 실재성 정도와는 별개로, 공신력 있는 기관에 의해서 기술 축적이나 기술개발에 관심을 가지고 있는 기업임을 인정받았다 할 수 있다. 이 외에도 기업의 여러 특성 중 연구개발 활동의 척도로서 활용할 수 있는 요소들은 최대한 반영하여 R&D 활동성 지수를 산출하고 있다.

다음의 〈표 2-1〉은 R&D 활동성 지수의 산출에 사용되는 다양한 항목을 정리한 표이다.

〈표 2-1〉 R&D 활동성 지수 산출을 위해 사용한 항목

분 류	항목명
기업 보유 특허	연도별 실용신안/특허 출원/등록 연도별 출원 특허 청구항 연도별 등록 특허 청구항 특허의 시장/권리/기술 점수 특허 출원 후 등록 비율 특허 출원 후 유지 비율 특허 출원 연속성
연구개발 비용	연도별 연구개발비 투입 추이 연도별 매출액 대비 연구개발비 연도별 국가 R&D 연구비 연구개발비 투입 연속성
연구개발 인력	연도별 연구인력 추이 R&D 인력 비율
국가연구개발사업 참여 여부	연도별 국가 R&D 연구 내용
기업 특성	연도별 연구개발 수행 여부 이노비즈/메인비즈/벤처기업 인증 연도별 연구기관 보유/운영 추이

자료 : 한국평가데이터에서 작성한 'R&D 활동성 지수 소개'를 토대로 저자 작성.

제 3 장
연구개발 활동의 생산성 및 고용효과 분석

제1절 기초통계량

본 장에서는 한국평가데이터에서 작성한 한국기업데이터와 R&D 활동성 지수를 이용하여 개별 사업체 단위의 패널자료를 구축한 후 실증 분석을 진행하였다. 본 절에서 관심을 가지는 주요 설명 변수는 R&D 활동성 지표(R&D 점수, R&D 환경 점수, R&D 투입 점수, R&D 산출 점수), 기업의 연구개발비 투자 금액, 특허 관련 정보이다. 이러한 정보를 이용하여 기업의 연구개발 활동이 생산성 및 고용에 미치는 영향을 다각도로 분석하고자 시도하였다.

1. R&D 활동성 지수

우선 한국기업데이터에서 제공하는 R&D 활동성 지수를 기준으로 기업의 산업분류 유형별로 R&D 수준에 대해서 평가한 결과를 요약한 것이 〈표 3-1〉이다.

첫째 특징으로 C26(전자부품, 컴퓨터, 영상, 음향 및 통신장비 제조업)과 C27(의료, 정밀, 광학 기기 및 시계 제조업)에서 R&D 활동성 기준 최우수 등급으로 분류된 사업체의 비율이 다른 산업보다 높게 나타났으며 미흡 판

정을 받은 사업체가 차지하는 비중은 가장 낮은 것으로 집계되었다. 반면, C10(식료품제조업), C22(고무 및 플라스틱제품 제조업), F(건설업), G(도소매업), J63(정보서비스업) 등에서는 R&D 활동성 기준으로 최우수 판정을 받은 사업체의 비율이 낮은 반면 미흡 판정을 받은 사업체의 비율은 높게 나타났다. 이러한 두 집단의 특성 차이는 해외 수출이 차지하는 비중과 해외 경쟁 사업자들의 국내 시장 진입의 수월성에서 기인하는 것으로 보인다. 예

〈표 3-1〉 산업별 R&D 수준에 대한 평가 결과

산업분류	미 흡		보 통		우 수		최우수		총 합
C10	10,450	75%	2,659	19%	669	5%	187	1%	13,965
C20	9,053	53%	5,456	32%	1,981	12%	654	4%	17,144
C22	9,027	67%	3,465	26%	813	6%	149	1%	13,454
C25	15,241	69%	5,687	26%	1,035	5%	180	1%	22,143
C26	9,757	41%	7,702	33%	4,336	18%	1,874	8%	23,669
C27	6,562	37%	5,922	34%	3,534	20%	1,598	9%	17,616
C28	10,645	48%	7,445	34%	3,074	14%	844	4%	22,008
C29	25,440	53%	15,530	33%	5,204	11%	1,512	3%	47,686
C30	7,321	60%	3,545	29%	1,126	9%	294	2%	12,286
C99	35,610	70%	11,283	22%	3,021	6%	984	2%	50,898
F	22,267	79%	4,874	17%	937	3%	155	1%	28,233
G	26,249	78%	6,068	18%	1,311	4%	239	1%	33,867
J58	15,976	42%	12,489	33%	7,329	19%	2,391	6%	38,185
J62	3,465	51%	2,266	33%	900	13%	191	3%	6,822
J63	2,391	59%	1,269	31%	379	9%	44	1%	4,083
J99	2,164	68%	820	26%	186	6%	13	0%	3,183
M	17,622	54%	9,715	30%	3,718	11%	1,336	4%	32,391

주 : 1) C10(식료품제조업), C20(화학물질 및 화학제품 제조업), C22(고무 및 플라스틱제품 제조업), C25(금속 가공제품 제조업), C26(전자부품, 컴퓨터, 영상, 음향 및 통신장비 제조업), C27(의료, 정밀, 광학 기기 및 시계 제조업), C28(전기장비 제조업), C29(기타 기계 및 장비 제조업), C30(자동차 및 트레일러 제조업), C99(기타 제조업), F(건설업), G(도소매업), J58(출판업), J62(컴퓨터 프로그래밍, 시스템 통합 및 관리업), J63(정보서비스업), J99(기타서비스업), M(전문, 과학 및 기술 서비스업)

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

를 들어, 건설업은 해외 건설 사업자가 국내에서 대규모 사업을 진행하기 어려울 뿐만 아니라 공공 건설은 국내 사업자들이 주로 수주하기 때문에 해외 사업자들의 국내 시장 진출도 쉽지 않다. 반면 전자부품이나 의료 및 정밀과학 기기는 해외의 경쟁력 있는 사업자들이 국내 시장에 이미 진출해 있으며 해외 사업자들의 제품 수준도 높은 편이기 때문에 국내 기업들도 일정 수준 이상의 제품 경쟁력을 갖추기 위해서는 품질 제고를 위한 연구개발 활동을 꾸준히 전개할 수밖에 없다.

둘째 특징은 대부분의 산업에서 R&D 활동성 기준으로 미흡 판정을 받은 기업의 비율이 가장 높고, 그다음으로 비율이 높은 등급이 보통 수준이며, 우수나 최우수 판정을 받은 기업들의 비율이 높지 않다. 특히 최우수 판정을 받은 기업들의 경우, 그 비율이 모두 10% 이하로 집계되어, 여전히 우리나라 기업들의 연구개발 활동이 아주 활발하지 않음을 확인할 수 있다.

R&D 활동성 지수의 평가 등급이 기업 규모별로 차이가 있는지를 살펴본 기초 통계 결과가 〈표 3-2〉이다. 이를 통해 보면, 우선 한국기업데이터와 R&D 활동성 지수 자료에 모두 등장하는 사업체의 기업 규모별 비율에 있어 중소기업이 대기업 및 중견기업보다 약 34배 정도 많다. 물론 이는 실제 대기업 및 중견기업의 수 대비 중소기업 수의 비율보다는 중소기업의 비율이 낮게 잡힌 수치이나 여전히 분석 데이터 내에서도 중소기업이 큰 비율을 차지함을 확인할 수 있다.

기업 규모별 집계 결과에 따르면 중소기업은 대기업 및 중견기업보다 미흡과 보통 수준의 판정을 받은 사업체의 비율이 높은 반면, 우수 및 최우수 판정을 받은 사업체의 비율은 대기업과 중견기업에서 높게 관측된다. 따라서 기업의 규모에 따라서 R&D 활동성의 정도가 다르다는 점을 확인할 수 있다.

다만, 앞선 〈표 3-1〉에서 나타난 R&D 활동성 지수상에서 미흡 및 보통 수준의 기업이 많은 것 역시 중소기업이 전체 표본에서 차지하는 비중이 크기 때문이라고 해석하는 것은 바람직하지 않다. 〈표 3-2〉의 결과에 따르면 심지어 대기업과 중견기업에서도 반 이상의 사업체가 미흡 판정을 받았으며, 보통 판정을 받은 기업까지 포함하면 대기업 및 중견기업의 75% 이상이 R&D 활동성 지수상에서 보통 혹은 그 이하의 R&D 활동성을 가지고 있는

〈표 3-2〉 기업 규모별 R&D 수준에 대한 평가 결과

R&D 수준	대기업/중견기업		중소기업	
미흡	5,858	51%	223,382	59%
보통	3,016	26%	103,179	27%
우수	1,762	15%	37,791	10%
최우수	854	7%	11,791	3%
총 합	11,490	100%	376,143	100%

주 : 기업 규모에 대한 분류는 한국기업데이터의 기업 분류 기준을 따랐음.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

것으로 나타났기 때문이다. 따라서 미흡 판정을 받은 기업들의 비율이 중소기업에서 더 높은 것은 사실이지만, 이러한 경향은 대기업과 중견기업에서도 공히 나타나기 때문에, 앞선 미흡 판정 기업들이 가장 많은 현상은 기업 규모의 분포에 따른 결과라기보다는 한국 기업들의 전반적인 경향이라 해석하는 것이 타당할 것이다.

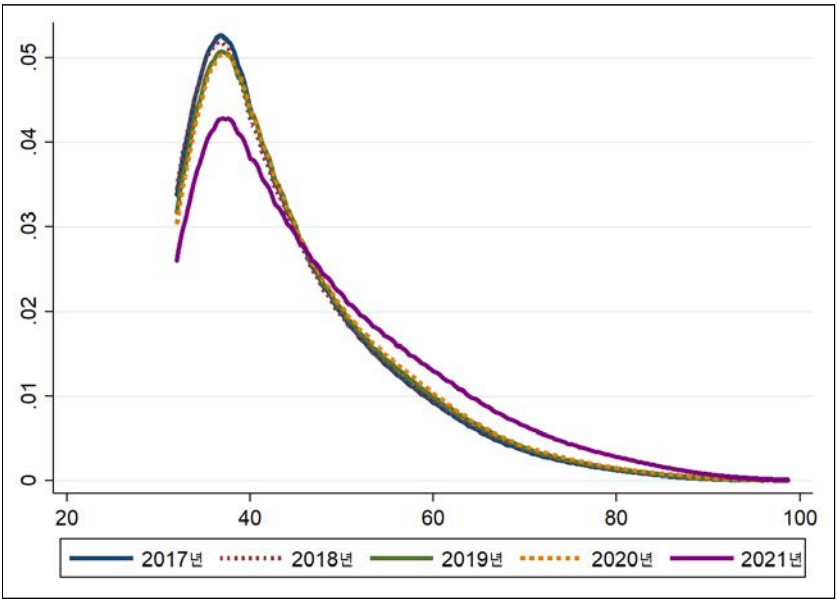
R&D 활동성 지수 자료는 2017년부터 2021년까지 제공받았다. 이를 바탕으로 0~100으로 매겨지는 R&D 활동성 지수 원점수의 연도별 분포를 그림으로 그린 것을 [그림 3-1]에서 확인할 수 있다.

2017년부터 2020년까지는 점수대별 사업체의 비율 분포가 거의 유사하지만, 2021년의 경우 낮은 점수대의 기업 비율은 줄어든 반면, 50점 이상을 기록한 사업체의 비율은 늘어난 것이 확인된다. 여기에는 두 가지 원인이 있을 수 있다.

첫째로, 코로나19의 영향이다. 그러나 코로나19는 2020년 초반부터 시작되었으므로 만일 코로나19가 기업의 연구개발 활동에서 변화를 초래했다면 이러한 경향은 2020년에도 관측되었어야 한다. 하지만 그렇지 않았으므로 분포 변화의 원인을 코로나19로 해석하는 것은 타당치 않을 것이다.

둘째로 표본 수에서의 차이이다. 2017년부터 2020년까지는 표본의 수가 모두 8만 개 사(社) 이상이었으므로 전체 표본의 구성이 유사하였을 가능성이 높다. 그러나 2021년에는 점수 집계에서의 시차 및 기업 재무정보의 입수와 정리 시기의 문제 등으로 인해 분석에서 가용한 기업 자료에서의 표본

[그림 3-1] 연도별 R&D 활동성 지수의 평가 점수 분포



주 : 1) 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 kernel density estimation을 추정한 결과임.
2) 연도별 표본 수 : 83,859(2017년) / 89,349(2018년) / 90,875(2019년) / 91,508(2020년) / 32,042(2021년)

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

의 수가 다른 연도보다 적게 관측되었다. 연도별 표본 수는 [그림 3-1]에 기술되어 있다. 그 결과 규모별 기업 분포에 있어 다른 해와 확연히 차이가 존재한다면 이러한 결과가 나타날 수 있다.

따라서 이후 기초 통계에 있어서 최근의 결과를 보여줄 때는 2021년 대신 2020년의 결과를 보여줄 것이다.

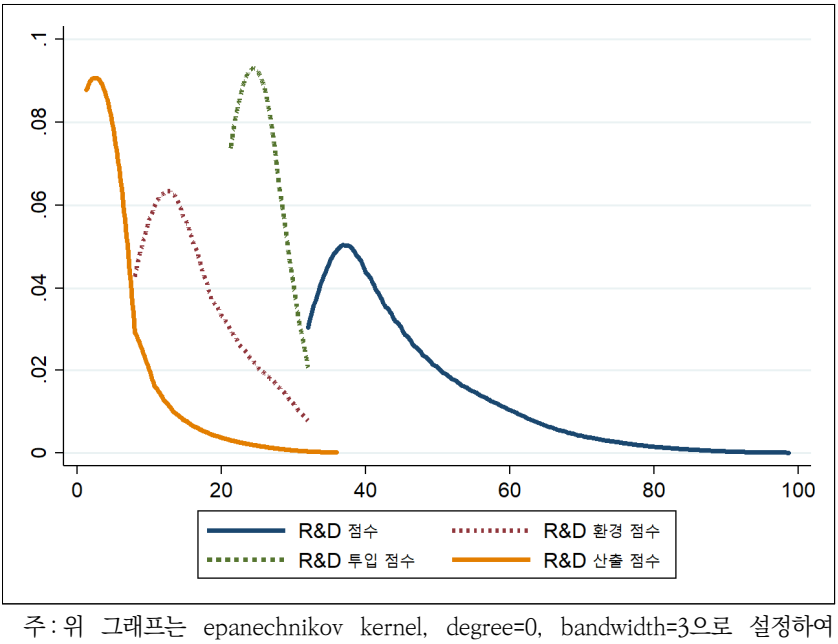
R&D 활동성 지수는 네 가지로 구성되어 있는데, 전체 점수, 환경점수, 투입점수, 그리고 산출점수이다. 이들 점수 각각이 서로 어떠한 관계를 가지고 있는지를 보기 위해서 개별 점수의 분포를 계산하였다.

R&D 환경점수는 해당 기업이 연구소를 보유하고 있는지, 연구개발비에는 얼마를 투자하는지 등 연구개발과 관련된 제반 환경을 평가하는 항목이다. R&D 투입점수는 연구개발에 투자하는 비용의 규모가 기업의 매출이나

자산 대비 어느 정도인지, 그리고 연구개발에 투입되는 인력은 전체 종사자 수와 비교하여 어느 정도인지에 따라서 매겨진다. 다시 말해서 R&D 투입점수는 연구개발 활동에 대한 기업의 재정적·인적 투입 정도를 측정한 지표라 할 수 있다. R&D 산출점수는 연구개발 활동의 결과물인 특허나 지적재산권의 양 및 그 우수성을 측정한 것이다. 여기서 우수성이란 특허나 지적재산권과 같은 연구개발 활동에서 나온 결과물의 가치가 높은지, 결과물의 산출이 단절 없이 지속적으로 이루어지고 있는지 등을 평가한 지표이다. 그리고 상술한 세 가지 지표인 환경점수, 투입점수, 산출점수를 근거로 기업의 R&D 활동성 정도를 수치화한 것이 R&D 활동성 지수의 원점수이다.

개별 점수 항목마다 모두 점수의 분포가 정적 편포(right-skewed) 되어 있음이 확인된다. 즉, 최빈치가 중위수보다 작고 중위수가 평균치보다 작아서 분포에서 오른쪽 꼬리가 길게 늘어진다는 것이다. 이는 상위 점수대에서

[그림 3-2] R&D 활동성 지수 평가 항목별 점수 분포(2020년)

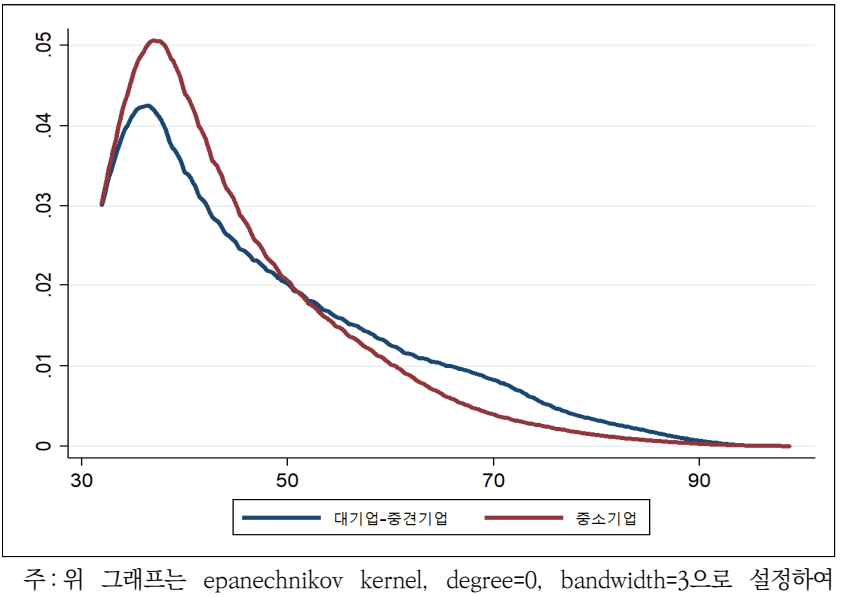


주 : 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 kernel density estimation을 추정한 결과임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

이상치(outlier)가 존재하거나 다수의 관측치가 낮은 값에 분포할 때 등장하는데, 여기서는 다수의 기업들이 R&D 활동성이 낮은 상태이기 때문에 나타난 결과로 보인다. 이는 〈표 3-1〉의 결과와도 유사한데, 대부분의 기업이 미흡 판정을 받는 것은 비단 원점수의 분포에 따른 결과만이 아니라 각 항목별로도 다수의 기업들이 낮은 점수를 기록하기 때문에 전체 점수에서도 유사한 분포가 그려지고 있음이 확인된다.

[그림 3-3]은 R&D 활동성 지수를 기업 규모에 따라 구분하여 대기업과 중견기업의 점수 분포와 중소기업의 점수 분포를 하나의 그림에 그린 것이다. 이 그림의 전반적인 분포는 〈표 3-2〉가 함의하는 바와 유사하다. 전체적으로 낮은 점수대에 기업들의 분포가 높게 나타나고 높은 점수대에는 소수의 기업만이 존재한다. 다만 중소기업 쪽에서는 낮은 점수대의 비율이 상대적으로 높고 대기업과 중견기업 쪽에서는 높은 점수대의 비율이 상대적으로 높다.

[그림 3-3] R&D 활동성 지수의 기업 규모별 점수 분포(2020년)



주 : 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 kernel density estimation을 추정한 결과임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

그러나 〈표 3-2〉와 비교하여 [그림 3-3]에서 특기할 만한 점은 대기업과 중견기업에서 최빈치의 점수가 더욱 낮다는 점이다. 점수의 차이가 크지는 않으나, 중소기업에서의 최빈치 점수보다 대기업과 중견기업에서의 최빈치 점수가 더 낮고, 대신에 최빈치를 기록한 기업 수의 비율은 대기업과 중견기업에서보다 중소기업이 더 높다.

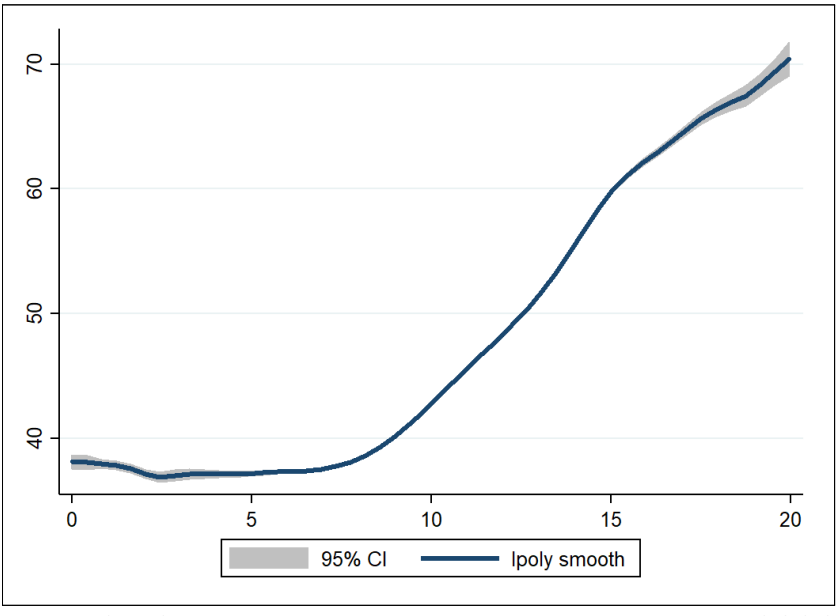
이를 통해 중소기업에서는 기업들의 R&D 활동성에서의 편차가 상대적으로 작다는 것을 알 수 있다. 상대적으로 대기업과 중견기업에서는 기업들의 점수대별 분포에서 중소기업보다 상대적으로 널리 퍼져 있는 비율을 보이는 반면, 중소기업의 경우에는 특정한 점수대에 반 이상의 기업이 몰려 있으며, 이후 점수가 증가하거나 감소함에 따라 급격하게 기업 수의 비율이 감소한다. 즉, 중소기업들은 대체로 비슷한 정도의 R&D 활동성을 보이고 있으며, 업종이나 중소기업 내 기업 규모, 지역 등에서 큰 차이를 보이고 있지 않다는 것이다.

그러나 대기업이나 중견기업의 경우, 업종이나 지역 혹은 집단 내에서의 기업 규모에 따라 R&D 활동성에서 차이를 보이고 있음을 알 수 있다. 기업 분포의 정적 편포가 대기업과 중견기업 집단 내에서 더 크게 나타나는 것을 통해 규모가 큰 기업이라 할지라도 연구개발이 필요하지 않은 업종에 있는 기업들은 오히려 중소기업보다 R&D 활동을 적게 하고 대신 규모의 경제에 따른 경쟁력 등을 가지고 시장에서 활동하는 것이 아닌가 하는 추측을 가능케 한다.

[그림 3-4]는 본 장에서 기업의 R&D 활동에 대한 주요 측정지표로 활용할 R&D 활동성 지수가 실제 기업이 지출하는 연구개발 투자 비용과 얼마나 밀접한 관계를 가지는지를 살펴봄으로써 해당 지표의 유효성을 간접적으로 살펴보고자 한다. 만일 R&D 활동성 지수가 연구개발 투자비와 전혀 상관관계를 가지지 않는 것으로 나타난다면 R&D 활동성 지수가 실제 기업의 연구개발 활동을 잘 대리하는 변수라는 가정이 성립하지 않기 때문이다.

가로축은 로그 변환한 기업의 연구개발 투자 비용이며 세로축은 R&D 활동성 지수이다. 두 변수 간의 국소다항회귀(local polynomial regression)를 수행한 결과는 두 변수가 양의 상관관계를 가지고 있음을 보여준다. 다만 로그 변환한 연구개발 투자비에서 8 이하의 값을 가지는 구간에서는 양자

[그림 3-4] R&D 활동성 지수와 연구개발 투자비 간의 관계(2020년)



주 : 1) 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 국소다항회귀분석을 수행한 결과임.
2) X축은 log(연구개발 투자비), Y축은 R&D 활동성 지수이며, 음영은 95% 신뢰구간임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

간의 상관관계가 전혀 관찰되지 않았음이 특기할 만하다.

이는 일정한 수준 이하의 연구개발 비용을 지출하는 경우에는 연구개발 활동에 있어서 좋은 성과물이 나오기 어렵거나, 일시적으로만 연구개발 활동을 수행하여 지속적으로 양질의 연구개발 활동 성과물을 내기에는 충분치 않거나, 혹은 연구개발 활동이 실질적이지 않을 가능성을 내포한다. 즉, 연구개발 활동에서도 일정한 규모 이상의 투자가 이루어져야지만 연구개발 활동이 제대로 된 성과를 낼 수 있음을 암시한다 하겠다.

R&D 활동성 지수가 기업이 수행하는 연구개발 활동에 대한 적절한 측정지표가 될 수 있음은 [그림 3-4]의 음영으로 표시된 95% 신뢰구간에서도 확인할 수 있다. 로그 변환한 수치 기준으로 8 이상의 구간에서 R&D 활동성 지수와 기업의 연구개발 투자 지출 간에는 양의 상관관계를 가질 뿐만 아니

라 95% 신뢰구간의 폭도 매우 좁게 나타나 양자 간의 양의 상관관계에 대한 추정 결과가 두 수치의 관계를 매우 잘 설명하고 있다.

결론적으로 [그림 3-4]는 일정 수준 이상의 R&D 활동을 전개하는 기업에 대해서는 R&D 활동성 지수가 기업의 연구개발 활동에 대한 적절한 측정지표임을 보여준다 할 수 있다.

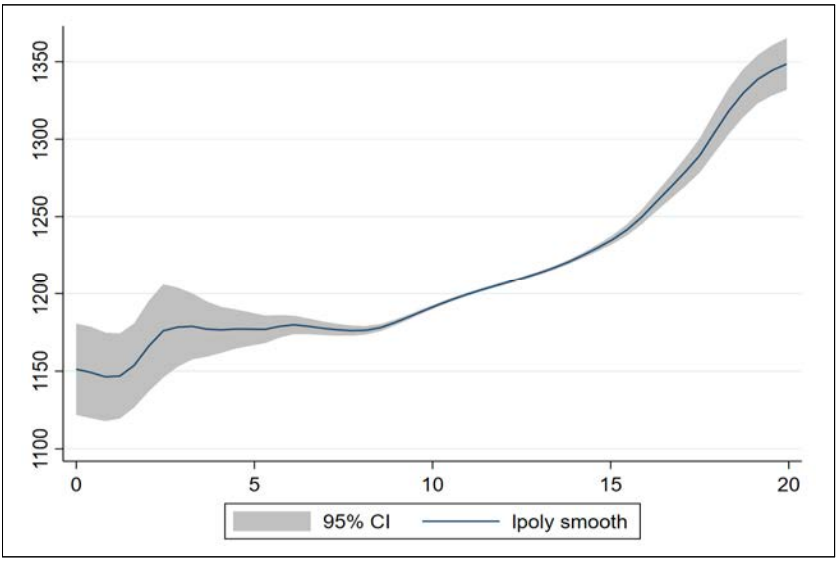
2. 기업의 연구개발 활동과 경영성과

본 장에서 기업의 생산성을 측정하기 위해 사용하는 지표는 종사자 1인당 매출액이다. 영업이익의 경우, 자료는 가용하나 음의 생산성이 측정되는 문제가 발생한다. 이를 수정하기 위해서 음의 영업이익을 0으로 환산할 수도 있으나, 이 경우 실제 0의 영업이익을 기록한 곳과 음의 영업이익을 기록한 곳이 생산성 측면에서 같게 측정된다는 또 다른 문제가 발생한다. 일반적으로 매출액과 영업이익 간의 상관관계가 높기 때문에 본 연구에서는 1인당 매출액을 생산성의 주요 측정지표로 삼을 것이다. 다만 여기서 기업의 연구개발 활동이 실제 매출액과 영업이익에 다른 방향의 영향을 주는지를 살펴보기 위해서 기초 통계에 있어서는 영업이익과 매출액을 모두 살펴볼 것이다.

현재 연구에서 가용한 연구개발 활동에 대한 측정지표는 기업이 지출한 연구개발에 대한 투자비용과 기업의 R&D 활동성 지수가 있다. 따라서 경영성과 지표에 대해서 두 가지 연구개발 활동과 관련된 수치들과 차례로 비교하고자 한다.

우선 [그림 3-5]는 1인당 매출액과 연구개발 투자비 간의 관계를 그린 것이다. 로그 변환한 연구개발 비용의 지출액과 로그 변환한 1인당 매출액을 비교해 보면, 전반적으로 양의 상관관계를 갖는 것이 확인된다. 다만, 앞서의 [그림 3-4]와 유사하게, 로그 변환한 연구개발 투자 비용의 수치가 8 이하인 구간에서는 음영으로 표시된 95% 신뢰구간의 범위가 넓게 나타나 추정결과에 대한 신뢰 수준이 낮을 뿐만 아니라, 1인당 매출액과 연구개발 투자비용 간에 통계적으로 유의한 상관관계가 관찰되지 않는다. 다시 말해서, 신뢰구간이 넓게 설정되고 그래프의 기울기가 0에 가까워짐을 의미한다. 그러

[그림 3-5] 1인당 매출액과 연구개발 투자비 간의 관계(2020년)



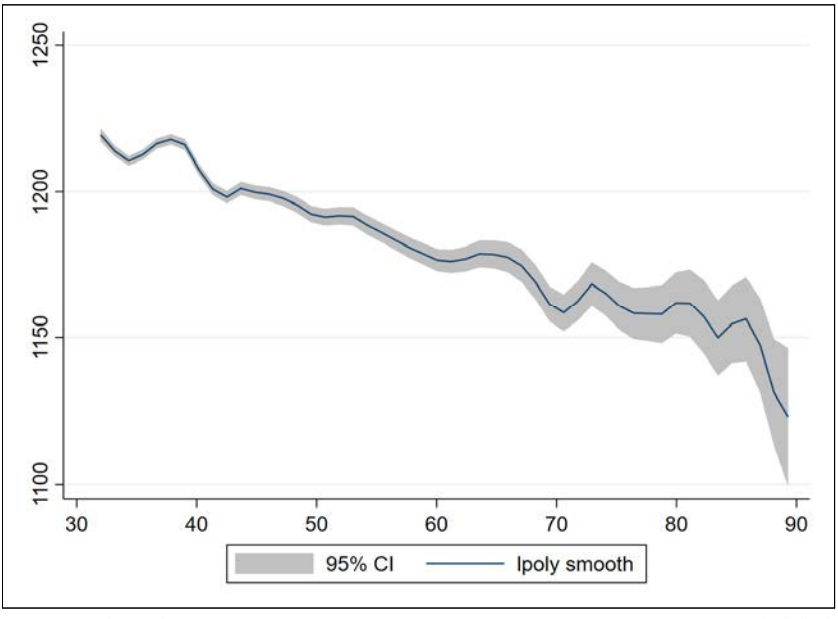
주 : 1) 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 국소다항회귀분석을 수행한 결과임.
2) X축은 log(연구개발 투자비), Y축은 1인당 매출액이며, 음영은 95% 신뢰구간임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

나 로그 변환한 연구개발 투자비가 8 이상인 구간에서는 두 변수 간에 동행성이 확인된다.

[그림 3-6]은 연구개발 활동의 또 다른 지표인 R&D 활동성 지수를 이용하여 1인당 매출액과 연구개발에 대한 투자 간의 관계에 대해서 동일하게 국소다항회귀분석한 결과이다. 해당 그림에서는 R&D 활동성 지수와 1인당 매출액이 서로 부(負)의 상관관계를 가지고 있음이 확인된다. R&D 활동성 지수가 60점 이하인 구간에서는 95% 신뢰구간도 좁게 추정되어 이러한 부의 관계가 더욱 뚜렷하게 드러난다. R&D 활동성 지수가 60점 이상인 기업의 경우, 상대적으로 신뢰구간의 폭이 넓기 때문에 R&D 활동성 지수가 아주 높은 기업들에서는 부의 관계가 명확하다고 말할 수는 없으나, R&D 활동성 지수가 늘어남에 따라 1인당 매출액이 줄어드는 경향성은 뚜렷이 존재한다고 할 수 있다.

[그림 3-6] 1인당 매출액과 R&D 활동성 지수 간의 관계(2020년)



주 : 1) 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 국소다항회귀분석을 수행한 결과임.
2) X축은 R&D 활동성 지수, Y축은 기업 단위 1인당 매출액이며, 음영은 95% 신뢰구간임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

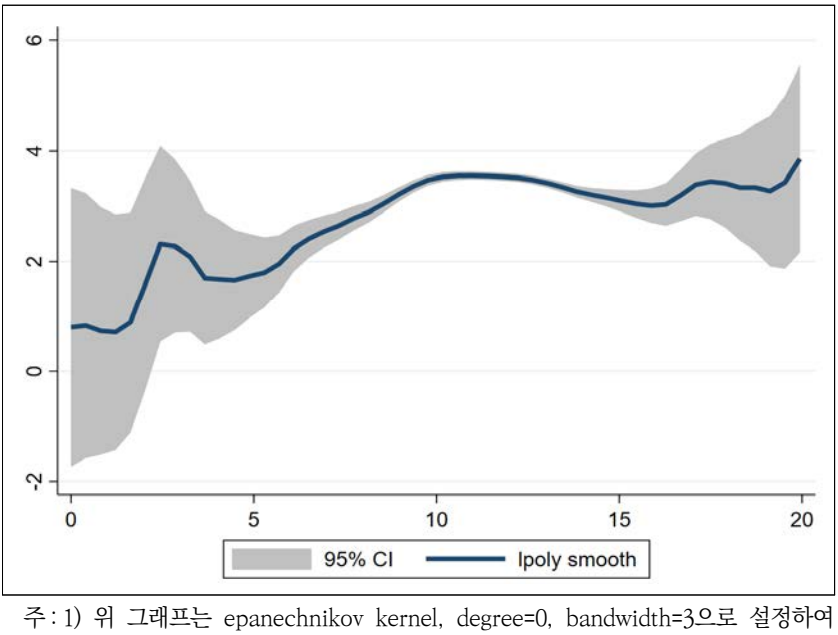
이러한 결과는 해석에서 두 가지의 해석 가능성을 제시한다. 첫째로 R&D 활동성 지수가 기업의 연구개발 활동에 대한 적절한 대리지표가 아닐 가능성이다. 그러나 [그림 3-4]에서 연구개발 투자비와 R&D 활동성 지수 간의 높은 양의 상관관계를 관찰할 수 있었기에 이러한 가능성은 비현실적이다. 둘째로는 [그림 3-4]와 연계하여 대부분의 기업이 연구개발 투자비 지출에 있어 로그 변환 기준 8 이하의 수치를 기록할 가능성이다. 특히나 표본에서 많은 크기를 차지하는 중소기업의 경우 R&D 활동성 지수도 낮을 뿐만 아니라 연구개발을 위한 지출도 적게 할 터인데, 이러한 구간에 포함된 다수의 기업들이 생산성 지표와 R&D 활동성 지수 간의 상관관계 결과를 왜곡한다는 것이다. 이는, R&D 활동성 지수가 낮거나 미흡한 구간에 위치한 기업들의 수가 가장 많으며, 해당 집단에서는 R&D 활동성 지수와 매출액 간에도

유의미한 관계가 관찰되지 않거나 혹은 부의 관계가 나타날 수 있는데, 이들 집단에서의 상관관계가 양 지표 간의 상관관계를 음으로 도출시켰을 수 있다.

추가적인 확인을 위해서 영업이익률과 R&D 활동성 지수 및 연구개발 투자비 간의 관계를 살펴보았다.

[그림 3-7]은 매출액 대비 영업이익의 비율인 영업이익률과 로그 변환한 연구개발 투자비 간의 관계를 국소다항회귀분석으로 추정한 것이다. 전반적으로는 양자 간에 양의 상관관계가 있는 것처럼 보이지만, 양극 구간 근처에서는 상대적으로 넓은 신뢰구간이 형성되어 통계적 유의성이 낮음을 알 수 있다. 특히 로그 변환한 연구개발 투자비의 값이 5 이하이거나 15 이상인 구간에서 이러한 현상이 강하게 관찰된다. 연구개발은 앞서도 언급하

[그림 3-7] 영업이익률과 연구개발 투자비 간의 관계(2020년)



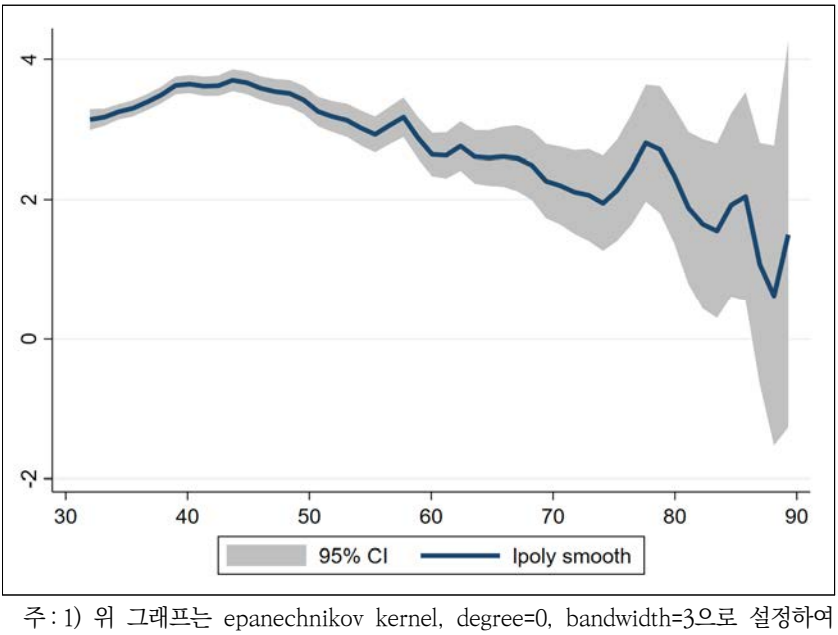
주 : 1) 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 국소다항회귀분석을 수행한 결과임.
2) X축은 log(연구개발 투자비), Y축은 영업이익률(영업이익/매출액)이며, 음영은 95% 신뢰구간임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

였듯이 불확실성이 높은 투자 활동이기 때문에 당장의 연구개발 비용 지출은 현 시점에서의 영업이익률을 낮출 수 있다. 다만 연구개발 활동이 즉각적인 효과를 내거나 혹은 꾸준한 연구개발 활동으로 지속적인 높은 영업이익을 창출하는 기업이 있을 수 있다. 또한 연구개발 활동을 전개하지 않고도 어느 정도의 영업이익을 확보할 수 있는 업종의 특성이나 사업체 고유의 특성으로 인해 연구개발 투자 지출이 낮은 집단에서 뚜렷한 상관관계가 나타나지 않을 수 있다.

그렇다면 R&D 활동성 지수는 기업의 영업이익률과 어떠한 관계를 가지는지도 점검할 필요가 있다. 그 결과가 [그림 3-8]이다. [그림 3-5]와 [그림 3-6] 간의 서로 상충하는 관계와 유사한 듯 하나, [그림 3-7]과 [그림 3-8] 간에는 공통점이 더 많다. [그림 3-7]과 비교하여 [그림 3-8]에서는 일부 구간

[그림 3-8] 영업이익률과 R&D 활동성 지수 간의 관계(2020년)



주 : 1) 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 국소다항회귀분석을 수행한 결과임.
2) X축은 R&D 활동성 지수, Y축은 영업이익률(영업이익/매출액)이며, 음영은 95% 신뢰구간임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

에서 음의 상관관계가 관찰은 되지만 일부 구간에서는 양의 상관관계가 나타나며, 또한 R&D 활동성 지수가 높은 구간에서는 오히려 영업이익률과 R&D 활동성 지수 간에 통계적으로 유의미한 관계를 전혀 관찰할 수 없었다.

특히 R&D 활동성 지수가 60 이상인 기업들에서는 영업이익률과 R&D 활동성 지수 간의 관계가 0으로 추정되는데, 이는 앞서 연구개발 투자비에서 연구개발 투자비 지출이 높은 기업 집단에서의 결과와 일치한다. 따라서 연구개발 투자비를 지속적으로 투입할 수밖에 없으며, 이로 인해서 영업이익에 일정 정도 손해를 보지만 경쟁이 치열하며, 또한 연구개발 활동의 성과에 따른 매출 변동이 큰 업종에 종사하는 기업들은 영업이익을 지속적으로 희생하면서도 연구개발 활동에 자금을 투입할 수밖에 없음을 추론할 수 있다.

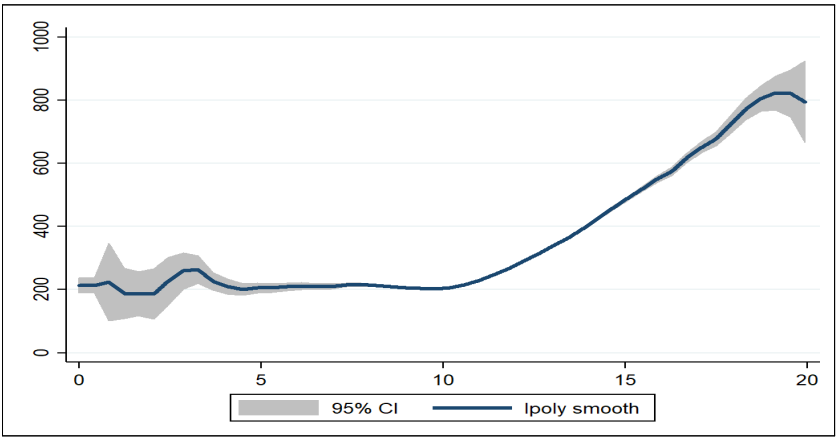
3. 기업의 연구개발 활동과 고용

한편 본 연구의 주요 관심 지표는 고용량에 있다. 따라서 고용과 기업의 연구개발 활동 간에는 어떠한 관계가 있는지를 국소다항회귀분석을 통해 살펴볼 필요가 있다.

[그림 3-9]는 국소다항회귀분석을 수행한 결과인데, 두 변수 간의 관계는 [그림 3-5]와 유사한 양상을 보여주고 있다. 로그 변환한 연구개발 투자비가 10 이하의 수치를 기록한 구간에서는 연구개발 투자비와 고용원 수 간에 특별한 상관관계를 찾기 힘들다. 그러나 연구개발 투자비가 아주 높은 일부 구간을 제외하면 그 수치가 10보다 큰 구간에서는 양자 간의 뚜렷한 양의 상관관계가 관찰된다. 그러므로 일정한 규모 이상의 기업은 연구개발 활동을 위한 투자를 활발하게 진행하고 있음이 확인된다.

고용원 수와 연구개발 활동 간에는 R&D 활동성 지수로 측정한 연구개발 활동의 활발한 정도가 고용원 수와 더욱 뚜렷하게 양의 상관관계를 보이는 것이 [그림 3-10]에서 확인된다. R&D 활동성 지수가 높은 구간에서는 95% 신뢰구간의 폭이 넓어지는 것이 확인되나, 이는 해당 점수대의 기업 수가 적기 때문으로 추정된다. [그림 3-10]을 통해 R&D 활동성 지수는 고용과 연구개발 활동 간의 관계를 분석하는 데 있어 유효한 자료일 수 있음을 확인하였다.

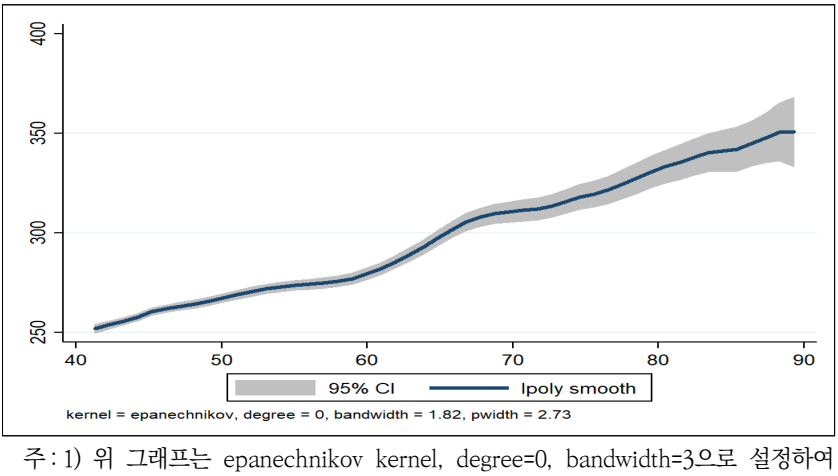
[그림 3-9] 고용원 수와 연구개발 투자비 간의 관계(2020년)



주 : 1) 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 국소다항회귀분석을 수행한 결과임.
2) X축은 연구개발 투자비, Y축은 로그 변환한 고용원 수이며, 음영은 95% 신뢰구간임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

[그림 3-10] 고용원 수와 R&D 활동성 지수 간의 관계(2020년)



주 : 1) 위 그래프는 epanechnikov kernel, degree=0, bandwidth=3으로 설정하여 국소다항회귀분석을 수행한 결과임.
2) X축은 R&D 활동성 지수, Y축은 로그 변환한 고용원 수이며, 음영은 95% 신뢰구간임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

제2절 기업의 R&D 투자가 생산성에 미치는 영향

지금부터는 구축한 기업 패널자료를 활용하여 기업, 업종, 지역 등 다양한 요인을 통제한 후, 기업의 연구개발 투자비 지출과 R&D 활동성 지수가 실제 기업의 생산성에 어떠한 방향으로 어느 정도의 영향을 미쳤는지를 추정하고자 한다.

1. 기업의 연구개발 투자비가 생산성에 미친 영향

우선 직접적인 연구개발 활동의 지표인 연구개발 투자비가 1인당 매출액으로 측정한 기업의 생산성에 어떠한 영향을 미쳤는지를 여기서 살펴보고자 한다. 산업 대분류 단위에서 고정효과를 통제하고 표본에 대해서 기업 규모를 고려한 가중치를 부여하지 않은 결과가 〈표 3-3〉이다. 그러나 일반적으로 연구개발 활동은 그 성과가 나오기까지 시차가 소요되므로, 1년의 시차를 둔 생산성 지표와 연구개발 투자비 간의 관계도 포함시켜 분석하였다. 한편 종사자 1인당 자본량(K/L)은 연구개발 투자 지출과 무관하게 1인당 매출액에 큰 영향을 미칠 수 있으므로 이를 추가적인 통제 변수로 투입하였다. 고정효과에 있어서는 그 통제 범위를 다양하게 설정하여 우리가 관심을 가지고 있는 연구개발 활동 관련 변수와 종속변수인 생산성 간의 관계를 최대한 강건하게 추정하고자 하였다.

우선 연구개발 투자비의 로그 변환치가 1인당 매출액으로 대리되는 종속변수인 생산성과 어떠한 관계를 가지는지 분석한 결과에 따르면, 연구개발 투자액이 로그 변환 기준 1 단위 증가하면 1인당 매출액도 약 3~4% 정도 증가하는 것으로 나타났다. 또한 연구개발 투자비와 생산성 간의 관계는 모든 모형에서 1% 유의수준하에 통계적으로 유의한 것으로 추정되어, 양자 간의 양의 상관관계가 뚜렷하게 관찰되었다.

더하여 1인당 자본량을 추가적으로 통제한 경우, 1인당 자본량은 기업의 생산성과 밀접하게 관련된 것으로 나타났으며, 그러나 여전히 연구개발 투

〈표 3-3〉 연구개발 투자비가 생산성에 미친 영향 : 산업 대분류 단위, 가중치 없음

$\beta \times 100$	(1)	(2)	(3)
log(연구개발 투자비)	3.527*** (.2316)	3.385*** (.3232)	2.701*** (.2067)
L.log(연구개발 투자비)		-.2489 (.2858)	
log(1인당 자본)(K/L)			30.52*** (.6070)
개별 기업 단위 고정효과	Y	Y	Y
연도×산업대분류×사업체규모	Y	Y	Y
연도×산업중분류×사업체규모	X	X	X
표본 가중치 설정 여부	X	X	X
관측치	259,185	165,175	246,205
조정 결정계수	.8286	.8523	.8534

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

자비의 계수값도 1% 신뢰수준에서 통계적으로 유의한 것으로 나왔다.

산업 대분류 단위에서는 이질적인 기업들이 동일한 산업으로 묶인 경우가 많다. 따라서 동일한 분석을 산업 중분류 단위의 고정효과를 부여하는 것으로 추가적으로 실시하였다. 그 결과는 〈표 3-4〉와 같다.

〈표 3-4〉에서 나타난 결과는 〈표 3-3〉과 대동소이하여 연구개발 투자비는 생산성과 양의 상관관계를 가지며 그 관계는 상당히 강건함을 확인할 수 있다.

여기서 사용한 종속변수는 로그 변환한 1인당 매출액인데, 가중치가 없는 모형에서는 매출액과 종사자 수가 많은 사업체와 그렇지 않은 사업체가 모두 동일하다는 전제하에 분석이 진행되었다. 그러나 규모가 큰 기업에 종사자 수가 많을 것이므로 연구개발 활동이 실제 경제의 생산성에 미치는 영향은 이러한 기업 간 규모 격차를 고려한 가중치를 부여하여 추가적인 분석을 수행함으로써 강건성을 검증해야 한다.

〈표 3-4〉 연구개발 투자비가 생산성에 미친 영향 : 산업 중분류 단위, 가중치 없음

$\beta \times 100$	(1)	(2)	(3)
log(연구개발 투자비)	3.461*** (.2309)	3.345*** (.3244)	2.662*** (.2065)
L.log(연구개발 투자비)		-.2966 (.2860)	
log(1인당 자본)(K/L)			30.35*** (.6050)
개별 기업 단위 고정효과	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y
표본 가중치 설정 여부	X	X	X
관측치	258,604	164,816	245,661
조정 결정계수	.8292	.8526	.8538

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

〈표 3-5〉는 산업 고정효과를 대분류 단위로 부여한 후 가중치를 사용하여 분석한 결과이다.

분석 결과는 〈표 3-3〉에서 가중치를 부여하지 않은 분석과 전반적으로 크게 다르지 않았다. 다만, 추정된 연구개발 비용의 효과에 대한 계수가 〈표 3-3〉과 비교하여 일관되게 줄어들었고, 대신 1인당 자본의 효과는 더욱 크게 나타났다. 이는 가중치가 매출액에 따라 부여된 것인데 이러한 매출액 규모가 1인당 자본량과 밀접한 변수이기 때문으로 보인다.

산업 고정효과를 중분류 단위에서 부여한 가중치를 포함한 분석 결과는 앞선 대분류 단위 고정효과 분석 결과와 크게 다르지 않다. 또한 모든 분석에서 시차를 둔 연구개발 투자비는 모두 유의하지 않은 것으로 나타나 연구개발 투자가 기업의 생산성에 미치는 효과는 일반적인 생각보다는 빠르게 나타날 수 있음을 시사한다.

〈표 3-5〉 연구개발 투자비가 생산성에 미친 영향 : 산업 대분류 단위, 가중치 부여

$\beta \times 100$	(1)	(2)	(3)
log(연구개발 투자비)	2.756*** (.4959)	2.694*** (.9199)	2.143*** (.2541)
L.log(연구개발 투자비)		.3741 (1.085)	
log(1인당 자본)(K/L)			45.42*** (2.240)
개별 기업 단위 고정효과	Y	Y	Y
연도×산업대분류×사업체규모	Y	Y	Y
연도×산업중분류×사업체규모	X	X	X
표본 가중치 설정 여부	Y	Y	Y
관측치	258,604	164,816	245,205
조정 결정계수	.8292	.8126	.8834

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

〈표 3-6〉 연구개발 투자비가 생산성에 미친 영향 : 산업 중분류 단위, 가중치 부여

$\beta \times 100$	(1)	(2)	(3)
log(연구개발 투자비)	2.696*** (.5292)	2.558*** (.9594)	2.182*** (.2434)
L.log(연구개발 투자비)		-.2471 (.9235)	
log(1인당 자본)(K/L)			42.19*** (1.177)
개별 기업 단위 고정효과	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y
표본 가중치 설정 여부	Y	Y	Y
관측치	258,604	164,816	245,205
조정 결정계수	.8207	.8157	.8859

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

2. 기업의 R&D 활동성 지수가 생산성에 미친 영향

지금까지 기업의 연구개발 투자비 지출액이 생산성에 어떠한 영향을 미쳤는지 살펴보았으며, 여기서는 기업의 연구개발 활동에 대한 종합적 지표인 R&D 활동성 지수가 생산성에 미친 영향이 어떠한지를 앞서와 유사하게 분석해 보았다.

R&D 활동성 지수가 1인당 매출액으로 측정한 생산성과 어떠한 관계를 가지는지 분석한 결과에서는, R&D 활동성 지수가 생산성에 미치는 영향이 모형마다 다른 것으로 나타났다.

예를 들어, 모형 (1)에서는 R&D 활동성 지수가 1점 증가할 때마다 1인당 매출액은 약 0.2% 정도 증가하는 것으로 추정되었다. 그러나 R&D 활동성 지수의 시차 효과까지 고려한 모형 (2)에서는 반대로 R&D 활동성 지수의 계수는 음수로 바뀌고 오히려 1년 전의 시차 변수 값이 양의 상관관계를 가지는 것으로 분석되었다. 그리고 1인당 자본량까지 통제한 모형 (3)에서는 R&D

〈표 3-7〉 R&D 활동성 지수가 생산성에 미친 영향 : 산업 대분류 단위, 가중치 없음

$\beta \times 100$	(1)	(2)	(3)
R&D 활동성 지수	.1765*** (.0410)	-.5370*** (.0578)	-.0499 (.0379)
전년도 R&D 활동성 지수		.4011*** (.0540)	
log(1인당 자본)(K/L)			32.23*** (.4842)
개별 기업 단위 고정효과	Y	Y	Y
연도×산업대분류×사업체규모	Y	Y	Y
연도×산업중분류×사업체규모	X	X	X
표본 가중치 설정 여부	X	X	X
관측치	375,954	252,986	351,870
조정 결정계수	.8027	.8290	.8335

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

활동성 지수는 생산성과 어떠한 상관관계도 가지지 않는 것으로 나타났다.

〈표 3-8〉은 산업별 고정효과를 대분류 단위가 아닌 중분류에 부여한 분석이다. 앞선 〈표 3-7〉과 유사하게 R&D 활동성 지수와 생산성 간의 관계는 모형에 따라 차이를 보이고 있다.

〈표 3-8〉 R&D 활동성 지수가 생산성에 미친 영향 : 산업 중분류 단위, 가중치 없음

$\beta \times 100$	(1)	(2)	(3)
R&D 활동성 지수	.1501*** (.0406)	-.5389*** (.0578)	-.0624 (.0375)
전년도 R&D 활동성 지수		.3802*** (.0539)	
log(1인당 자본)(K/L)			32.19*** (.4832)
개별 기업 단위 고정효과	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y
표본 가중치 설정 여부	Y	Y	Y
관측치	375,954	252,432	351,097
조정 결정계수	.8032	.8293	.8339

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

당해연도 R&D 활동성 지수만을 고려한 모형에서는 R&D 활동성 지수와 생산성 간에 양의 상관관계가 확인되었으나, 이러한 관계는 전년도 R&D 활동성 지수나 1인당 자본과 같은 변수를 통제하면 상관관계가 사라지거나 혹은 음의 상관관계로 변하는 것으로 나타났다. 이를 통해 다음과 같은 사실을 확인할 수 있다.

첫째로, 당해연도와 직전년도 R&D 활동성 지수 간에는 높은 상관관계가 존재한다는 것이다. 그래서 전년도 R&D 활동성 지수를 고려하지 않은 모형에서는 전년도 R&D 활동성 지수의 효과와 당해연도 R&D 활동성 지수의 효과가 섞여 순효과가 양의 값으로 관찰된 것이다. 그러나 생산성에 더 큰 영

향을 미치는 전년도 R&D 활동성 지수를 통제하면 이러한 전년도 지수의 양의 효과가 제외된 금년도 지수의 효과가 음으로 나타났다.

둘째로 R&D 활동성 지수와 1인당 자본 간에도 높은 상관관계를 가지는 것으로 보인다. 그래서 모형 (1)에서는 1인당 자본과 연관된 생산성과의 양의 상관관계 일부가 R&D 활동성 지수에 잡혀서 계수가 양의 값으로 나왔을 것이다. 그러나 모형 (3)에서 이러한 1인당 자본의 효과를 통제한 경우 R&D 활동성 지수와 1인당 매출액 간에는 뚜렷한 상관관계를 발견하지 못하고 있다. 특히나 1인당 자본량을 모형에 넣으면 조정 결정계수가 크게 커지는 것이 확인되어 1인당 자본이 1인당 매출액과 큰 관계를 가지며 R&D 활동성 지수 그 자체는 기업의 생산성과 아주 강한 관계를 가진다고는 할 수 없다.

기업 규모를 고려한 가중치를 부여한 분석 결과가 〈표 3-9〉이다.

〈표 3-9〉 R&D 활동성 지수가 생산성에 미친 영향 : 산업 대분류 단위, 가중치 부여

$\beta \times 100$	(1)	(2)	(3)
R&D 활동성 지수	-.6487*** (.2070)	-1.216*** (.3589)	-.5476*** (.0570)
전년도 R&D 활동성 지수		.3099** (.1484)	
log(1인당 자본)(K/L)			46.08*** (1.715)
개별 기업 단위 고정효과	Y	Y	Y
연도×산업대분류×사업체규모	Y	Y	Y
연도×산업중분류×사업체규모	X	X	X
표본 가중치 설정 여부	Y	Y	Y
관측치	375,954	252,986	351,870
조정 결정계수	.8069	.8186	.8740

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

분석 결과에 따르면 R&D 활동성 지수는 일관되게 1인당 매출액으로 측정한 기업의 생산성과 부의 관계를 가지는 것으로 나타났다. 또한 지난 기의 R&D 활동성 지수와 1인당 자본은 모든 다른 분석에서처럼 생산성과 양의 관계를 가지는 것으로 나타났다. 특히나 1인당 자본을 포함시킨 모형에서 조정 결정계수 값이 가장 높았는데, R&D 활동성 지수가 1점 높아질 때마다 생산성이 약 0.55% 정도 감소한다는 결과를 얻었다.

이는 [그림 3-4]와 연결시켜 해석할 필요가 있다. 즉, 연구개발 투자를 거의 하지 않는 기업 집단에서는 생산성과 R&D 활동성 지수 간의 부의 관계 혹은 상관관계가 없음이 나타나며 R&D 활동성 지수 기준으로 미흡 판정을 받은 기업이 전체 기업의 반 이상이므로, 이들까지 포함한 전체 결과에서는 이러한 R&D 활동에서의 하위 집단으로 인해 음의 관계가 도출되었을 가능성이 높다.

〈표 3-10〉 R&D 활동성 지수가 생산성에 미친 영향 : 산업 중분류 단위, 가중치 부여

$\beta \times 100$	(1)	(2)	(3)
R&D 활동성 지수	-.5745*** (.2075)	-1.208*** (.3839)	-.4967*** (.0466)
전년도 R&D 활동성 지수		.3478** (.1200)	
log(1인당 자본)(K/L)			43.72*** (1.021)
개별 기업 단위 고정효과	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y
표본 가중치 설정 여부	Y	Y	Y
관측치	375,954	252,432	351,870
조정 결정계수	.8069	.8227	.8763

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

산업 고정효과를 중분류 단위에서 부여하고 기업의 1인당 매출액을 기준으로 한 가중치를 포함한 분석 결과도 역시 가중치를 고려한 대분류 결과와 크게 다르지 않은 R&D 활동성 지수 및 생산성 간의 부의 관계를 보여주었다.

3. 강건성 검증을 위한 기업 규모별 연구개발 투자비와 생산성의 관계 분석

R&D 활동성 지수를 이용한 분석과 연구개발 투자비를 이용한 분석 간 결과의 괴리를 검증하기 위해서 기업 규모별로 연구개발 투자비와 생산성 간의 관계를 추가적으로 살펴보았다. 그 결과는 〈표 3-11〉에 정리되어 있다.

〈표 3-11〉 기업 규모별 연구개발 투자비가 생산성에 미친 영향 : 산업 중분류 단위, 가중치 부여

$\beta \times 100$	대기업 및 중견기업		중소기업	
	(1)	(2)	(3)	(4)
log(연구개발 투자비)	2.131 (1.325)	2.218** (.9319)	2.740*** (.5613)	2.177*** (.2508)
log(1인당 자본)(K/L)		56.38*** (6.561)		40.89*** (1.032)
개별 기업 단위 고정효과	Y	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y	Y
표본 가중치 설정 여부	Y	Y	Y	Y
관측치	8,523	8,305	351,870	237,356
조정 결정계수	.9005	.9364	.7846	.8617

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

대기업 및 중견기업만을 대상으로 한 모형 (1) 및 (2)와 중소기업만을 분석한 모형 (3) 및 (4) 간의 결과에서 뚜렷한 차이를 발견할 수는 없었다. 다만 몇 가지 특징을 꼽자면, 첫째로 대기업 및 중견기업 집단을 대상으로 한 분석에서 1인당 자본과 생산성 간의 상관관계가 중소기업보다 더 크게 나타났다. 이는 규모가 큰 기업일수록 규모의 경제에 의한 생산성 증대 효과가 컸거나, 혹은 업종의 특성상 규모의 경제 효과가 크게 발현되는 곳에서 대기업이 많았으며 그 결과 규모의 경제 효과도 더욱 크게 나타날 수 있다는 점이다.

둘째로 조정 결정계수를 보면 대기업과 중견기업에서 중소기업보다 더 높은 결정계수 값이 나왔음이 확인 가능하다. 따라서 연구개발비와 생산성 간의 관계는 기업 규모가 클수록 분석에 유효한 표본이 많거나 혹은 실제 연구개발을 수행하는 기업들이 전체 표본에서 차지하는 비중이 커지기 때문에 분석 결과의 설명력이 높을 수 있음을 시사한다.

마지막으로 연구개발 투자비와 생산성 간의 직접적인 양의 상관관계는 중소기업에서 더욱 높은 유의수준하에 나타난다는 것이다. 이는 중소기업 집단 내에서는 연구개발비를 지출하는지 여부가 실제 생산성의 격차와 밀접하게 연관되어 있음을 시사한다.

제3절 기업의 R&D 투자가 고용에 미치는 영향

앞서서는 기업의 연구개발 투자비 지출과 R&D 활동성 지수가 기업의 생산성에 미친 영향을 분석하였다. 여기서는 이러한 지표들이 기업의 고용과 어떠한 관계를 가지고 있는지를 알아보고자 한다.

1. 기업의 연구개발 투자비가 고용에 미친 영향

〈표 3-12〉는 업종 대분류 단위에서 연구개발 투자비와 고용의 관계를 분석한 결과이다. 이때 고용원 수는 로그 변환한 값을 사용하였다. 고용원 수

는 일반적으로 기업의 규모 혹은 총매출과도 밀접하게 관련되어 있기 때문에 분석에서 로그 변환한 매출액 및 로그 변환한 전기 매출액을 통제하였다.

한편 고용원 수는 기업이 단기에 조정하기 힘들기 때문에 연구개발 활동을 활발히 전개한다 하더라도 즉각적으로 바뀌기 어려울 수 있다. 또한 고용원 수에 있어서 기업 고유의 특성이 영향을 미치는 내생성이 존재할 가능성도 있다. 그러므로 이러한 내생성과 고용의 경직성을 통제하기 위해서 전기의 고용원 수를 포함한 분석도 추가로 실시해 보았다.

〈표 3-12〉 연구개발 투자비가 고용에 미친 영향 : 산업 대분류 단위, 가중치 없음

$\beta \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
log(연구개발 투자비)	5.113*** (.1313)	4.351*** (.1667)			2.496*** (.1307)	
L.log(연구개발 투자비)		4.167*** (.1641)	3.871*** (.1511)	2.625*** (.1367)		2.032*** (.1302)
log(매출액)					10.56*** (.3338)	
log(전기 매출액)				13.97*** (.3711)		11.81*** (.3587)
log(전기 고용원 수)					.2182*** (.0071)	.1876*** (.0077)
개별 기업 단위 고정효과	Y	Y	Y	Y	Y	Y
연도×산업대분류×사업체규모	Y	Y	Y	Y	Y	Y
연도×산업중분류×사업체규모	X	X	X	X	X	X
표본 가중치 설정 여부	X	X	X	X	X	X
관측치	262,218	166,396	182,162	183,931	178,385	183,008
조정 결정계수	.9566	.9684	.9658	.9690	.9703	.9702

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

분석 결과를 순서대로 해석해 보면, 우선 모형 (1)에서는 연구개발 투자비도 로그 변환한 값이고 종속변수인 고용원 수 역시 로그 변환한 값이므로 이는 탄력성을 의미한다고 볼 수 있다. 즉, 연구개발 투자비가 단위 비율만큼 변하면 고용원의 수는 어느 정도의 비율로 변화하는지를 의미하는 것이다. 모형의 (1)의 결과에 따르면 연구개발 투자비가 100% 증가한다면 고용원 수는 5.1% 증가한다.

(2)열의 결과는 연구개발 투자비에 있어서 전기의 수치까지 고려하여 연구개발 활동에 따른 시차도 통제한 모형이다. 이에 따르면 연구개발 투자비의 시차 변수도 고용원 수와 밀접하게 연관된 것으로 보인다. 또한 전기와 당해연도 기의 연구개발 투자비가 모두 비슷한 정도의 상관관계를 보이고 있음이 확인된다.

모형 (3)에서는 연구개발 투자비와 고용 간에 상호 영향을 줄 수 있는 내생성의 문제를 보다 잘 통제하기 위해서 당해연도의 연구개발 투자비는 분석에 포함시키지 않고 전년도 연구개발 투자비만을 가지고 분석을 실시하였다. 내생성 문제로 인한 편의가 발생하는 경우에는 설명 변수에 외생적 요인에 의한 변수를 포함시키거나 혹은 선결 변수(pre-determined variable)를 포함시키는 것이 해결 방법이다. 추정 결과에 의하면 연구개발 투자비를 100% 증가시키면 다음 해의 고용원 수는 약 3.8% 증가하는 것으로 나타났다.

모형 (4)는 기업의 규모를 통제하기 위한 변수로 매출액을 분석에 포함시킨 결과이다. 이는 연구개발 활동과 무관하면서도 사업체의 규모를 통제할 수 있는 변수를 분석에 포함시킴으로써 기업의 규모 효과가 관심을 가지고 있는 변수의 효과로 잡히는 것을 막기 위한 방법이다. 그러나 매출액과 고용원 수는 밀접한 관계를 가지고 있을 가능성이 높으므로, 이러한 내생성을 통제하기 위해서 매출액에 대해서는 선결 변수를 모형에 포함시켰다. 그 결과 매출액을 통한 사업체의 규모를 통제한 경우, 연구개발 투자비가 100% 증가하면 고용원은 약 2.6% 증가하는 것으로 추정되었다.

한편 연구개발 투자비가 기업의 고용을 변화시키는 데는 보다 긴 시차가 필요할 수 있다. 따라서 이러한 동학(dynamics)을 고려하기 위해 설명변수로서 종속변수의 전기 값을 포함시키는 AR(1) 동태패널모형을 활용하여 모

형을 추정하였다. 그 결과 전기 고용원 수 역시 모두 유의한 변수로 나타났으고, 전년도 및 당해연도 연구개발 투자비의 로그 변환 값의 계수 역시 여전히 통계적으로 유의한 비슷한 값을 보였다. 보다 정확히는 (6)열 기준으로 전년도 연구개발 투자비가 100% 증가하는 경우 고용은 약 2% 증가한다고 할 수 있다.

AR(1) 모형의 계수값을 활용하여 연구개발 투자비가 증가할 때 고용에 미치는 장기적인 효과를 계산하면, 2.032/(1-0.1876)=2.501으로, 연구개발 투자비가 100% 증가할 경우 고용은 장기적으로 약 2.5% 증가한다고 할 수 있다.

〈표 3-13〉 연구개발 투자비가 고용에 미친 영향 : 산업 중분류 단위, 가중치 없음

$\beta \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
log(연구개발 투자비)	4.993*** (.1292)	4.295*** (.1667)			2.492*** (.1308)	
L.log(연구개발 투자비)		4.082*** (.1627)	3.803*** (.1498)	2.628*** (.1367)		2.047*** (.1303)
log(매출액)					10.48*** (.3342)	
log(전기 매출액)				13.77*** (.3715)		11.69*** (.3587)
log(전기 고용원 수)					.2144*** (.0070)	.1847*** (.0076)
개별 기업 단위 고정효과	Y	Y	Y	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y	Y	Y	Y
표본 가중치 설정 여부	X	X	X	X	X	X
관측치	261,622	166,033	185,755	183,529	178,011	182,618
조정 결정계수	.9572	.9687	.9661	.9692	.9704	.9703

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

업종에 대한 고정효과를 대분류 대신 중분류 단위로 보다 세분화하여 부여한 분석 결과에서도 결과의 방향성은 〈표 3-13〉과 〈표 3-12〉 간에 큰 차이를 보이지는 않았다. 다만, 업종 분류를 보다 세분화한 이후 전기 고용원과 당해연도 고용원 간 관계의 계수값이 확연히 감소하였다. 그러나 당해연도 연구개발 투자비를 비롯한 나머지 변수들의 계수값은 거의 변화가 없었으며, 따라서 업종을 중분류로 세분화하였을 때도 연구개발 투자비가 고용에 미치는 효과는 대분류 단위일 때와 큰 차이가 없이 같은 크기라고 해석해도 무방하다.

〈표 3-14〉 연구개발 투자비가 고용에 미친 영향 : 산업 대분류 단위, 가중치 부여

$\beta \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
log(연구개발 투자비)	5.167*** (1.180)	5.053*** (.7085)			2.918*** (.7784)	
L.log(연구개발 투자비)		4.823*** (.8733)	4.523*** (.7464)	2.954*** (.8085)		2.800*** (.7633)
log(매출액)					16.02*** (1.669)	
log(전기 매출액)				20.67*** (1.474)		19.66*** (1.430)
log(전기 고용원 수)					.0868 (.0673)	.0592 (.0644)
개별 기업 단위 고정효과	Y	Y	Y	Y	Y	Y
연도×산업대분류×사업체규모	Y	Y	Y	Y	Y	Y
연도×산업중분류×사업체규모	X	X	X	X	X	X
표본 가중치 설정 여부	Y	Y	Y	Y	Y	Y
관측치	14,786,198	10,799,267	11,304,633	11,281,106	11,211,555	11,267,777
조정 결정계수	.9931	.9954	.9953	.9955	.9953	.9955

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

가중치를 부여한 경우, 계수값의 크기는 거의 모든 변수에서 더 크게 나타났다. 그러나 한 가지 특기할 만한 사실은 전기 고용원 수에 대한 계수값이 모두 유의하지 않게 나타난 점이다. 가중치는 기업의 고용 인원에 따라 부여된 것이므로 앞서 나타난 고용원 수 증가 효과는 대부분 중소기업에서 집중되어 나타나며 규모가 큰 기업에서는 고용 증대 효과가 상대적으로 작다는 것을 시사한다. 다만, 연구개발 투자비가 고용에 대해 갖는 상관관계는 선결 변수이건 당해연도 변수이건 무관하게 앞선 두 표와 유사한 수준으로 나타났다.

〈표 3-15〉 연구개발 투자비가 고용에 미친 영향 : 산업 중분류 단위, 가중치 부여

$\beta \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
log(연구개발 투자비)	4.927*** (1.155)	4.734*** (.6694)			2.595*** (.7253)	
L.log(연구개발 투자비)		5.009*** (.8156)	4.697*** (.6944)	3.156*** (.7529)		3.001** (.7419)
log(매출액)					15.87*** (1.654)	
log(전기 매출액)				20.68*** (1.383)		19.79*** (1.288)
log(전기 고용원 수)					.0836 (.0644)	.0535 (.0602)
개별 기업 단위 고정효과	Y	Y	Y	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y	Y	Y	Y
표본 가중치 설정 여부	Y	Y	Y	Y	Y	Y
관측치	14,681,427	10,719,378	11,222,343	11,198,898	11,131,245	11,267,777
조정 결정계수	.9934	.9958	.9956	.9959	.9957	.9959

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

업종을 중분류 단위로 구분하고 가중치를 적용한 분석에서도 역시 〈표 3-14〉와 유사한 결과를 얻었으며, 가중치를 부여한 후에는 고용원 수에서의 내생성을 통제하기 위해 포함시킨 전기의 고용원 수가 통계적으로 유의하지 않게 나타났다.

분석에서 사업체별 고용인원을 가중치로 사용하는 경우 추정 계수값은 사업체의 서로 다른 고용인원을 추가적으로 고려하여 전체적인 고용 증가분에 대해 추정한 것으로 이해할 수 있다. 예를 들어, 어떠한 외생적 충격 혹은 정책 변화로 인해 100명을 고용하고 있는 사업체에서 고용을 10% 증가시키고 10명을 고용하고 있는 사업체에서 고용을 10% 감소시키는 경우, 사업체별 고용원 수를 이용한 가중치를 부여하지 않은 경우에는 모형의 추정 결과가 10% 증가와 10% 감소의 합인 불변으로 나타난다. 그러나 사업체별 고용원 수를 가중치로 활용한 모형에서는 전체 고용원의 증감이 +10명과 −1명으로 나타나 순효과를 +9명으로 추정한다. 따라서 고용원 수에 대한 분석에 있어서는 기업의 규모에 대해서 가중치를 부여하는 모형이 합리적이다. 또한 업종 구분 역시 대분류보다는 중분류 단위가 보다 세세하기 때문에 중분류 단위로 고정효과를 부여하는 것이 보다 합리적이다. 그러므로 〈표 3-12〉부터 〈표 3-15〉까지의 결과 중에서는 〈표 3-15〉가 가장 강건한 모형의 결과라고 판단할 수 있다.

2. 기업의 R&D 활동성 지수가 고용에 미친 영향

이제는 기업의 연구개발 활동에 대한 지표를 연구개발 투자비에서 R&D 활동성 지수로 바꾼 후 동일한 분석을 수행하여 그 결과를 살펴보겠다.

분석 결과에 따르면 R&D 활동성 지수가 1% 증가할 때마다 고용원 수는 0.7~1.33% 정도 증가하는 것으로 나타났다. 연구개발 활동의 시차 효과를 고려하여 이전 기의 R&D 활동성 지수를 포함시켜 분석한 결과에서도 현재 시점과 이전 기의 R&D 활동성 지수 모두 고용원 수와 양의 상관관계를 가진다는 점이 확인되었다.

또한 기업 규모를 통제하기 위해 이전 기나 당해연도의 매출액을 추가한 경우나 내생성을 통제하기 위해 전기 고용원 수를 추가한 경우에도 모두 당

〈표 3-16〉 R&D 활동성 지수가 고용에 미친 영향 : 산업 대분류 단위, 가중치 없음

$\beta \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
R&D 활동성 지수	1.329*** (.0215)	.7541*** (.0285)			.6853*** (.0243)	
전기 R&D 활동성 지수		.5033*** (.0258)	.8605*** (.0250)	.6179*** (.0234)		.4713*** (.0221)
log(매출액)					11.05*** (.3003)	
log(전기 매출액)				14.84*** (.3268)		13.10*** (.3214)
log(전기 고용원 수)					.1939*** (.0060)	.1636*** (.0063)
개별 기업 단위 고정효과	Y	Y	Y	Y	Y	Y
연도×산업대분류×사업체규모	Y	Y	Y	Y	Y	Y
연도×산업중분류×사업체규모	X	X	X	X	X	X
표본 가중치 설정 여부	X	X	X	X	X	X
관측치	380,235	254,940	254,940	252,041	251,588	250,822
조정 결정계수	.9490	.9608	.9603	.9643	.9649	.9654

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

해연도의 고용원과 R&D 활동성 지수 간에 양의 상관관계를 가짐이 확인되었다.

업종에 대한 고정효과를 중분류 단위로 부여한 분석에서도 전반적인 계수값의 크기나 부호는 대분류 단위 분석과 큰 차이를 보이지 않았다. 전체적으로 계수값들의 크기가 조금씩 줄어든 경향은 발견되었으나, 모든 변수에 대해서 그러한 하락 추세가 관찰된 것은 아니다.

기업의 규모에 따른 가중치를 부여하지 않은 분석에서는 앞서 언급한 바와 같이 개별 기업의 고용 변화율을 동일하게 처리하므로 총고용에서의 변화를 정확하게 측정하기 어렵다. 따라서 이를 위해 기업 규모에 대한 가중

〈표 3-17〉 R&D 활동성 지수가 고용에 미친 영향 : 산업 중분류 단위, 가중치 없음

$\beta \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
R&D 활동성 지수	1.294*** (.0213)	.7456*** (.0294)			.6827*** (.0243)	
전기 R&D 활동성 지수		.4851*** (.0258)	.8372*** (.0249)	.6154*** (.0234)		.4719*** (.0221)
log(매출액)					10.92*** (.2994)	
log(전기 매출액)				14.70*** (.3251)		13.01*** (.3190)
log(전기 고용원 수)					.1918*** (.0060)	.1621*** (.0062)
개별 기업 단위 고정효과	Y	Y	Y	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y	Y	Y	Y
표본 가중치 설정 여부	X	X	X	X	X	X
관측치	379,367	254,379	254,379	251,488	251,053	250,287
조정 결정계수	.9493	.9609	.9605	.9643	.9649	.9655

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

치를 부여하고 대분류 단위의 업종에 고정효과를 부여한 분석의 결과인 〈표 3-18〉을 보면, 앞선 〈표 3-16〉이나 〈표 3-17〉과 비교하여 R&D 활동성 지수의 전기 값, 매출액, 전기 매출액에서는 계수값의 크기가 더 커진 것을 확인할 수 있다.

그러나 R&D 활동성 지수는 계수값의 크기가 줄어들어, 가중치를 반영한 분석에서는 R&D 활동성 지수와 당해연도 고용원 수 간의 상관관계가 보다 낮아지는 것을 파악할 수 있다. 한편, 전기 고용원 수는 가중치를 고려한 경우 이전의 연구개발 투자비와 동일하게 아예 통계적 유의성을 상실한 것으로 나타났다.

〈표 3-18〉 R&D 활동성 지수가 고용에 미친 영향 : 산업 대분류 단위, 가중치 부여

$\beta \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
R&D 활동성 지수	.6479*** (.1721)	.2962*** (.1247)			.4959*** (.1348)	
전기 R&D 활동성 지수		.4951*** (.1665)	.6461*** (.1523)	.4862*** (.1601)		.4792*** (.7633)
log(매출액)					16.44*** (1.662)	
log(전기 매출액)				22.03*** (1.774)		21.02*** (1.729)
log(전기 고용원 수)					.0828 (.0521)	.0594 (.0506)
개별 기업 단위 고정효과	Y	Y	Y	Y	Y	Y
연도×산업대분류×사업체규모	Y	Y	Y	Y	Y	Y
연도×산업중분류×사업체규모	X	X	X	X	X	X
표본 가중치 설정여부	Y	Y	Y	Y	Y	Y
관측치	18,588,830	13,864,888	13,864,888	13,834,033	13,826,713	13,819,460
조정 결정계수	.9916	.9942	.9942	.9947	.9946	.9948

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

한편 가중치와 함께 업종별 고정효과를 중분류 단위에서 부여한 분석의 결과가 〈표 3-19〉이다. 〈표 3-16〉부터 〈표 3-19〉까지를 놓고 보면 강건성 측면에서 〈표 3-19〉의 결과가 가장 신뢰할 수 있다. 해당 결과를 놓고 보면 R&D 활동성 지수가 한 단위 증가할 때마다 고용원 수는 약 0.5~0.6% 증가하는 것이 확인된다. 따라서 규모가 크고 종사자 수가 많은 기업일수록 R&D 활동성 지수로 측정한 연구개발 활동이 활발하게 이루어지고 있다는 결론을 내릴 수 있다.

한편 설명변수의 내생성을 통제하고 사업체의 규모 효과를 매출액 변수를 통해 통제한 (4)열의 결과에 의하면 R&D 활동성 지수가 1점 높아지는 경

〈표 3-19〉 R&D 활동성 지수가 고용에 미친 영향 : 산업 중분류 단위, 가중치 부여

$\beta \times 100$	(1)	(2)	(3)	(4)	(5)	(6)
R&D 활동성 지수	.6442*** (.1716)	.3281*** (.1249)			.5252*** (.1305)	
전기 R&D 활동성 지수		.4779*** (.1589)	.6444*** (.1458)	.5046*** (.1519)		.4969*** (.1677)
log(매출액)					15.67*** (1.604)	
log(전기 매출액)				21.00*** (1.727)		20.16*** (1.674)
log(전기 고용원 수)					.0727 (.0495)	.0515 (.0484)
개별 기업 단위 고정효과	Y	Y	Y	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y	Y	Y	Y
표본 가중치 설정 여부	Y	Y	Y	Y	Y	Y
관측치	18,464,835	13,772,351	13,772,351	13,741,587	13,734,405	13,727,152
조정 결정계수	.9919	.9946	.9956	.9951	.9950	.9951

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

우 고용원 수는 약 0.5% 증가하는 것으로 추정되었다. R&D 활동성 지수의 변화가 고용원 수에 미치는 장기적인 효과를 고려하는 (6)열의 결과에 의하면 R&D 활동성 지수가 1점 상승하는 경우 고용인원은 1년 후에 약 0.49% 정도 증가한다. 장기적인 고용효과를 고려하기 위해서는 AR(1) 계수값을 활용해야 하나 (6)열에서 추정된 전기의 고용원 수에 대한 계수값이 통계적 유의성이 없으므로, 이 값을 이용하여 장기적인 효과의 크기를 계산하는 것은 적절하지 않다고 할 수 있다.

3. 강건성 검증을 위한 기업 규모별 고용효과 분석

본 소절에서는 R&D 활동성 지수와 연구개발 투자비 간의 불일치 문제가 발생하지는 않았다. 그러나 분석의 강건성을 재확인하기 위함과 동시에 기업 규모에 따라 고용효과에 차이가 있는지를 살펴보기 위해서 연구개발 지출비와 R&D 활동성 지수에 대해서 각각 규모별 고용효과를 추정하였다.

기업 규모별로 연구개발 투자비가 고용원 수에 미친 영향을 분석한 〈표 3-20〉에서 두 집단 간에 차이가 존재하는지를 보기 위해서는 (1)과 (3)에서의 로그 변환 연구개발 투자비의 계수가 서로 다른지, (2)와 (4)에서의 계수가 서로 다른지를 비교해야 한다.

〈표 3-20〉 기업 규모별 연구개발 투자비가 고용에 미친 영향 : 산업 중분류 단위, 가중치 부여

$\beta \times 100$	대기업 및 중견기업		중소기업	
	(1)	(2)	(3)	(4)
log(연구개발 투자비)	5.399*** (1.805)	4.117** (1.970)	4.284*** (.2657)	2.607*** (.2278)
	(1) - (3) 1.115 (1.824)		(2) - (4) 1.510 (1.984)	
log(매출액)		56.38*** (6.561)		40.89*** (1.032)
개별 기업 단위 고정효과	Y	Y	Y	Y
연도×산업대분류사업체규모	X	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y	Y
표본 가중치 설정 여부	Y	Y	Y	Y
관측치	11,222,343	11,198,898	11,222,343	11,198,898
조정 결정계수	.9956	.9959	.9956	.9959

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

(1)과 (3)의 경우 연구개발 투자비가 100% 증가하는 경우 고용원 수에 있어서 대기업과 중견기업에서는 약 5.4% 증가하고 중소기업에서는 4.3% 증가하는데, 두 값의 표준편차를 고려하면 두 집단 간의 차이인 1.1%는 통계적으로 유의하지 않은 것으로 나타난다. 또한 기업의 규모를 통제하기 위해 매출액까지 고려한 분석인 (2)와 (4) 간에도 통계적으로 유의한 차이를 찾지 못하였다.

따라서 중소기업과 대기업 및 중견기업 두 집단 간에 연구개발 투자비에 따른 고용효과에 있어서 차이가 존재한다고 결론내릴 수 없음이 확인되었다.

한편 R&D 활동성 지수의 고용효과를 측정하여 기업 규모별로 비교한 분석에서도 대기업 및 중견기업과 중소기업 간에 통계적으로 유의미한 격차

〈표 3-21〉 기업 규모별 R&D 활동성 지수가 고용에 미친 영향 : 산업 중분류 단위, 가중치 부여

$\beta \times 100$	대기업 및 중견기업		중소기업	
	(1)	(2)	(3)	(4)
log(연구개발 투자비)	.6531* (.3707)	.5873 (.3797)	.6389*** (.0590)	.4446*** (.0572)
	(1) - (3) 0.0142 (0.3754)		(2) - (4) 0.2538 (0.3824)	
log(매출액)		17.14*** (4.216)		23.07*** (.6122)
개별 기업 단위 고정효과	Y	Y	Y	Y
연도×산업대분류×사업체규모	X	X	X	X
연도×산업중분류×사업체규모	Y	Y	Y	Y
표본 가중치 설정여부	Y	Y	Y	Y
관측치	13,772,351	11,198,898	13,772,351	11,198,898
조정 결정계수	.9946	.9959	.9946	.9959

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

를 발견할 수 없음이 〈표 3-21〉을 통해 확인 가능하다. 그러므로 적어도 연구개발 활동에 따른 고용효과 측면에서는 기업 규모에 따른 유의미한 차이가 존재하지 않음을 확인할 수 있다.

다만, 이는 고용 변화율에서 유의미한 차이가 없음을 의미하며, 만일 고용량에 대한 논의라면 이야기가 다를 수 있다. 앞서 설명하였듯이 똑같이 10%가 변화했다 하더라도 대기업에서 전체 고용원의 10%가 변한 것과 중소기업에서 10%가 변한 것은 절대적 양에서는 차이가 크기 때문이다.

제4절 소 결

기업의 연구개발 활동은 즉각적이든 장기적이든 기업의 생산성에 영향을 미친다. 그리고 기업의 생산성은 매출액이나 기업 규모를 변화시킴에 따라 고용을 결정하는 주요한 변수로 작용한다. 따라서 기업이 전개하는 연구개발 활동이 실제로 기업의 생산성을 늘리는 효과가 있는지, 그리고 그 결과 고용은 어떻게 변하는지 살펴보는 것은 이후 연구개발 활동 및 특허 정책에 대해서 논의할 때 꼭 필요한 기초 자료라 할 수 있다.

본 장에서는 한국기업데이터의 연구개발 투자비와 한국평가정보에서 작성한 R&D 활동성 지수 두 가지 지표를 기업의 연구개발 활동을 측정하는 지표로 놓고 이 두 변수가 생산성 및 고용원 수와 어떠한 관계를 가지는지 살펴보았다. 그 결과는 다음과 같다.

첫째, 일정한 수준 이하의 연구개발 활동은 어떠한 지표로 관측하여도 유의미한 연구개발 활동으로 포착되지 않는다는 점이다. 다시 말해서 기업의 연구개발 활동이 유의미한 성과를 낼 확률을 높이거나 혹은 생산성 및 고용효과를 가지기 위해서는 일정 수준이나 규모 이상의 연구개발을 지속적으로 수행해야 한다는 것이 확인되었다.

둘째로, 기업의 연구개발 활동과 생산성 간의 관계에 있어서는 연구개발 투자비와 R&D 활동성 지수 간에 서로 상이한 결과를 보여주었으므로 단일한 결론을 도출하는 것이 불가능하였다. 연구개발 투자비와 1인당 매출액으

로 측정한 생산성 간에는 양의 상관관계가 뚜렷하게 관찰되었다. 그러나 R&D 활동성 지수와 생산성 간에는 음의 관계가 나타났다. 이를 경제학적 맥락에서 해석해 보자면, 인과관계가 아니라 상관관계임에 주목하여, 기업들이 연구개발 활동을 전개하고 이를 위해 투자비를 지출한다 하여도 1~2년 내에는 반드시 기업의 경영성과가 개선되는 것은 아니라는 점이다. 이러한 이유로는 첫째로 연구개발 활동은 성공이나 산출물이 보장되는 기업 활동이 아니기 때문이며, 둘째로 연구개발 활동의 산출물이 경영성과로 빠르게 이어지는 것이 아니고 셋째로 연구개발을 수행하기 위한 비용은 기업의 수익성을 단기적으로 혹은 중기적으로 악화시키는 지출 항목에 해당하기 때문으로 생각된다.

셋째, 생산성과 연구개발 활동에서의 결과와는 달리, 고용은 두 가지 연구개발 활동 지표 모두와 강한 양의 상관관계를 가지는 것을 확인하였다. 이것은 연구개발 활동을 전개하기 위해서는 생산 및 관리 인력뿐만 아니라 연구개발을 위한 인력이 추가로 필요하기 때문이기도 하며, 한편으로는 일정한 규모 이상의 기업이 효과적인 연구개발 활동을 전개할 수 있다는 역의 상관관계도 암시한다. 즉, 규모가 큰 기업이 연구개발 활동에 대한 실질적인 필요를 느끼며 그로 인해서 실질적인 연구개발 활동을 수행할 가능성이 있음을 시사한다.

제 4 장
국가연구개발사업의 생산성 및 고용효과

제1절 기초통계량

본 장에서는 2010년 이후 정부가 추진했던 국가연구개발사업이 연구개발 활동의 성과물이나 기업의 경영성과, 그리고 고용에 미친 영향을 추정하고자 한다. 첫째, 국가연구개발사업에서 목적했던 연구개발 투자가 특허로 측정되는 성과물에 미치는 영향에 대해 양적 분석을 시행할 것이다. 특히, 본 장에서는 전통적인 선형회귀분석의 한계점을 극복하기 위해 사건연구방법을 사용하여 국가연구개발사업의 수년간에 걸친 중장기적인 효과까지 추정하고자 한다.

둘째, 최근에 다양하게 활용되고 있는 준실험 인과관계 방법론을 활용하여 국가연구개발사업의 효과를 시각화하여 제시해 결과의 해석을 단순하고 투명하게 할 수 있도록 할 것이다. 사건연구방법을 활용하면 국가연구개발사업의 정책효과를 좀 더 정밀하게 추정할 수 있고 시각화에 기반하여 효과검증을 할 수 있다. 또한, 최근 국가 연구지원사업에서 선택과 집중의 대상이 되는 산업의 이질적인 효과에 대해서도 살펴볼 것이다.

국가연구개발 지원사업이 기업의 연구개발 성과에 미친 정책효과를 분석하기 위해서는 연구개발의 성과물을 수량화(quantify)할 필요가 있다. 기존의 선행 연구에서 대표적으로 사용되는 정량지표는 다음과 같다.

(ⅰ) 특허 수 : 회사가 취득한 특허의 수는 기술적인 혁신과 발견의 정도를 나타내는 지표

(ⅱ) 특허 출원 비율 : 특허 출원 수를 연구개발에 대한 지출로 나눈 비율로 연구개발 지출의 효율성을 측정하고 미래의 신기술 도입에 대한 정보를 포함.

(ⅲ) 특허 인용 횟수 : 회사의 특허가 다른 기업이나 연구자에 의해 얼마나 자주 인용되는지를 나타내는 지표로, 각 특허의 기술적인 영향력과 신기술의 질을 측정하는 것이 가능

(ⅳ) 신제품 출시 수 : 기업 입장에서 기술개발의 결과물로 새롭게 제작한 신제품이나 서비스의 출시 수는 기술 혁신에 대한 지표로 활용할 수 있음.

본 장에서는 학계에서 가장 많이 활용되고 있는 특허 개발 여부 및 한 해 동안 기업이 취득한 특허의 수를 성과변수로 사용할 것이다.

분석에서는 앞서 제2장에서 설명한 기업 단위 패널자료를 활용할 것이다. 구축된 패널데이터를 활용하면 지역 및 산업 단위 그리고 시점 단위의 세밀한 연구가 가능하나 여기서는 노동시장 성과지표나 기업 경영성과의 전체적인 모습을 살펴보는 데 집중하고 이질적 효과에 대한 심층적 분석은 추후 연구로 남겨 둔다.

〈표 4-1〉은 분석 표본에서 국가연구개발 지원사업에 참여한 기업과 그렇지 않은 기업을 구분하여 기초통계량을 비교하였는데, 관심 있는 결과변수인 특허 출원 여부를 비롯하여 기업의 고용량과 경영성과 변수를 확인할 수 있다. 국가연구개발 지원사업에 참여한 기업의 노동자 수는 평균 200명으로 한 번도 참여해 보지 않은 기업의 평균 노동자 수 17.6명보다 압도적으로 더 많다. 자산과 매출 등으로 측정한 기업 규모에 있어서도 국가연구개발 지원사업에 참여한 기업이 그렇지 않은 기업보다 훨씬 크게 나타났다.

〈표 4-1〉의 기초통계량으로 미루어 보아 커다란 표본 선택(Sample Selection) 문제가 존재하는 것으로 보인다. 따라서 기본적인 최소자승법에 의한 회귀분석 결과는 정책효과를 추정하는 데 적절하지 못할 것으로 예상되며, 국가연구개발 지원사업의 효과에 대한 인과관계 추정을 위해서는 적

〈표 4-1〉 분석 표본의 기초통계량(2010~2021)

	변수명	평균	표준편차
비교군 : 국가연구개발사업 참여 기업	특허 출원 여부	0.008	0.09
	고용원 수	17.682	13718.97
	로그 변환 매출	13.894	1.73
	로그 변환 자산	13.596	1.835
	로그 변환 총자본	12.716	1.829
	로그 변환 당기순이익	11.026	1.613
처리군 : 국가연구개발사업 미참여 기업	특허 출원 여부	0.226	0.418
	고용원 수	200.413	2634.902
	로그 변환 매출	15.042	2.348
	로그 변환 자산	15.288	2.186
	로그 변환 총자본	14.489	2.296
	로그 변환 당기순이익	12.288	2.199

자료 : 한국기업데이터 및 국가 과학기술 지식정보를 활용하여 저자 작성.

절한 방법론의 선택이 아주 중요하다. 그러므로 본 연구에서는 준실험 방법론 중 하나인 사건연구를 적용하여 정책효과 분석을 진행하고자 한다.

제2절 이중차분법에 따른 생산성 및 고용효과 분석 결과

이중차분법을 이용한 본 절에서는 다음의 모형을 사용하였다.

$$y_{i,t} = \beta_0 + \beta_1 treat_i + \beta_2 DiD_{i,t} + X_{i,t}\gamma + \mu_t + \epsilon_{i,t} \quad (1)$$

여기서 i는 기업을 t는 연도를 나타내며, $DiD_{i,t}$는 정책효과를 추정하는 더미변수로 여기서는 해당 t연도에 기업 i가 국가연구개발사업에 참여한 이력이 있는가 여부이다. μ_t는 연도별 고정효과를 통제하기 위한 변수이며, $X_{i,t}$는 기업의 시간에 따른 크기에 대한 정보를 포함하는 변수가 포함되었

다. $treat_i$는 기업 i가 한 번이라도 국가연구개발사업에 참여한 적이 있는지를 통제하기 위한 변수로, 참여 기업이 미참여 기업과 비교하여 가질 수 있는 특성 차이를 포착하기 위한 변수이다.

〈표 4-2〉는 (1)식을 추정한 이중차분법의 결과이다. 앞서 언급한 표본 선택의 문제를 줄이기 위해 여기서는 관측 기간 중 한 번이라도 국가연구개발사업의 지원을 받은 기업만으로 표본을 한정해서 추정하였다. 따라서 동일한 기업군에 대해서 국가연구개발사업 참여 전후의 정책효과를 비교하는 형태의 분석을 시행한 것이다. 종속변수와 자산 및 자본의 경우 절대치가 중요한 것이 아니기 때문에 분석에서는 모두 로그 변환한 값을 사용하였다.

추정 결과 국가연구개발사업에 참여한 기업은 사업 참여 이후 그전보다 특허 출원 수가 늘어나고 고용 규모와 자본 지출이 증가하는 것으로 나타났다. 보다 구체적으로, 특허 출원 수는 7% 증가하였으며, 노동자 수와 자본 사용은 각각 10%와 9% 증가하는 것으로 나타났다. 정책효과에 대한 추정치 모두 0.01의 유의수준에서 통계적으로 유의하게 나타났다. 그러나 당기순이익과 매출 성과는 오히려 21%와 17% 감소하는 것으로 나타났다. 다만,

〈표 4-2〉 이중차분법 수행 결과

	(1)	(2)	(3)	(4)	(5)
	특허 수	당기순이익	매출	고용원 수	납입자본금
정책 후 수혜 기업 (DID효과)	0.0658*** (0.00324)	-0.211*** (0.00960)	-0.174*** (0.00817)	0.100*** (0.00856)	0.0899*** (0.0104)
자산	0.0218*** (0.00174)	0.256*** (0.00778)	0.829*** (0.00615)	0.516*** (0.00587)	0.393*** (0.00717)
총자본	0.00994*** (0.00161)	0.568*** (0.00749)	0.109*** (0.00555)	0.150*** (0.00512)	0.278*** (0.00662)
상수항	-0.364*** (0.0164)	0.221*** (0.0414)	1.237*** (0.0363)	-5.169*** (0.159)	3.203*** (0.0495)
관측치	250912	204466	244364	207312	250264
결정 계수	0.0321	0.629	0.772	0.675	0.625

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.

자료 : 한국평가데이터에서 작성한 한국기업데이터 및 R&D 활동성 지수 자료를 이용하여 저자 작성.

(1)에 의한 추정 결과는 샘플 선택 문제를 완전히 해결하지 못한다. 따라서 이를 통제하기 위해 기업의 고정효과를 통제한 고정효과 모형을 가지고 추가적인 추정을 시행하였다.

기업 고정효과를 포함한 식은 다음과 같다.

$$y_{i,t} = \beta_0 + \beta_1 treat_i + \beta_2 DiD_{i,t} + X_{i,t}\gamma + c_i + \mu_t + \epsilon_{i,t}$$

기업 고정효과 모형 추정식은 c_i가 앞선 이중차분법의 식 (1)에 추가된 형태이다. 하지만 기업 고정효과를 추가함으로 인해 정책 참여 여부에 따른 기업의 특성을 통제하는 $treat_i$와 c_i 간에 높은 상관관계에서 기인한 다중공선성 문제 때문에 β_1은 추정될 수 없다. 따라서 기업 고정효과 모형 식의 경우는 정책효과를 추정하는 데 있어 정책이나 사업에 참여한 기업과 참여치 않은 기업 간 비교를 배제하고 기업의 연도 간 비교만을 사용하여 정책효과를 추정한다.

〈표 4-3〉에서는 기업 고정효과를 추가적으로 통제한 결과를 보여주고 있다. 〈표 4-2〉의 추정에서와 동일하게 표본 선택 문제를 통제하기 위해서 관측 기간 중 한 번이라도 국가연구개발사업의 지원을 받은 기업으로만 추정 대상을 한정하였다. 추정 결과 국가연구개발 지원사업에 참여한 기업은 참여 이후에 그전보다 특허 출원 수가 늘어나고 고용원 수와 자본 지출이 증가하는 것으로 나타났다. 이는 방향성 측면에 있어서 〈표 4-2〉의 결과와 동일하다. 그러나 효과의 강도 측면에서는 차이가 있다. 특허 출원 수는 4% 증가하는 것으로 나타났으며, 고용원 수와 자본 사용은 각각 8%와 5% 증가하는 것으로 분석되었다. 추정치들은 모두 여전히 통계적으로 유의하게 나타났다.

그러나 주요 경영성과 지표 중 하나인 당기순이익은 여전히 통계적으로 유의하게 7% 감소하는 것으로 나왔다. 반면 매출액은 증가하지도 감소하지도 않고 변화가 아주 작게 나타나 통계적으로 유의미한 변화를 찾을 수 없었다.

〈표 4-3〉 기업 고정효과 모형 기반 이중차분법 수행 결과

	(1)	(2)	(3)	(4)	(5)
	특허 수	당기순이익	매출	고용원 수	납입자본금
정책 후 수혜 기업 (DID효과)	0.0398*** (0.00327)	-0.0713*** (0.0101)	0.00754 (0.00575)	0.0870*** (0.00568)	0.0483*** (0.00474)
자산	0.0451*** (0.00209)	0.318*** (0.0109)	0.687*** (0.00698)	0.426*** (0.00628)	0.215*** (0.00614)
총자본	0.0109*** (0.00166)	0.506*** (0.0101)	0.114*** (0.00487)	0.0403*** (0.00392)	0.279*** (0.00601)
상수항	-0.673*** (0.0239)	0.291** (0.102)	3.091*** (0.0754)	-2.630*** (0.198)	5.634*** (0.0648)
관측치	250912	204466	244364	207312	250264
결정계수	0.288	0.751	0.915	0.892	0.926

주 : 1) *** p<0.01, ** p<0.05, * p<0.1.
2) (괄호)안의 값은 기업 단위로 군집화된(clustered) 표준오차임.
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 국가과학기술지식정보를 활용하여 저자 작성

제3절 사건연구 모형에 따른 생산성 및 고용효과 분석 결과

앞서 살펴본 기업 고정효과만으로는 충분치 않아 이번에는 정책 전후 효과를 매년 구분해서 보는 사건연구 모형을 적용하여 추정하였다.

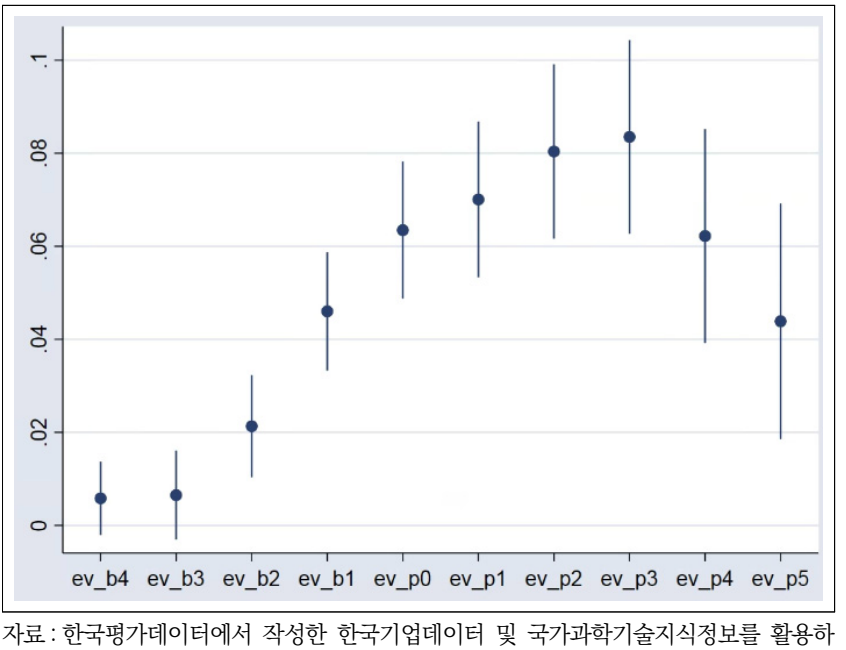
$$y_{i,t} = \beta_0 + \Sigma_{k \in [-4,-3,\ldots,0,1,\ldots,5]} \beta_k \tau(k)_{i,t} + X_{i,t}\gamma + c_i + \mu_t + \epsilon_{i,t}$$

여기서, $\tau(k)_{i,t}$는 주어진 표본에서 기업 i가 최초로 국가연구개발 지원사업에 참여한 연도와의 거리를 나타내는 더미변수이다. 예를 들어, 이벤트 연도(k=0)인 최초로 국가연구개발 지원사업에 참여한 연도가 2002년이고 현재 연도 t가 2004년이면 $\tau(k=2)_{i,t}$=1이고 나머지 k에 대한 $\tau(k)_{i,t}$은 0이 된다. 종속변수는 이전과 같이 로그 변환을 하여 추정하였다.

[그림 4-1]은 사건연구 모형으로 추정하여 -4부터 5까지 각각의 k에 대해서, β_k의 계수 추정치와 그 값의 95% 신뢰구간을 표시하였다. 추정 결과에 따르면, 정책 수혜를 받기 전까지 특허 수는 계속 증가하는 추세에 있었으며, 국가연구개발사업의 지원을 받는 순간부터 이후 3년째까지는 특허 수가 이전과 같은 추세로 증가했다. 그러나 4년 이후에는 더 이상 증가하지 않고 오히려 감소하는 것으로 나타났다. 단순 사건 전후 평균을 비교하면 국가연구개발 지원사업 수혜 이후에 특허 출원 수가 많았지만 지원사업 참여 이전의 추세까지 고려하면 사업 참여 이전부터 존재하던 기존의 증가 추세가 이어지지는 못하고 있다.

기업의 연구개발 활동은 즉각적이든 장기적이든 기업의 생산성에 영향을 미친다. 그리고 기업의 생산성은 매출액이나 기업 규모를 변화시킴에 따라 고용을 결정하는 주요한 변수로 작용한다. 따라서 기업이 전개하는 연구개발 활동이 실제로 기업의 생산성을 늘리는 효과가 있는지, 그리고 그 결과

[그림 4-1] 사건연구 모형 추정 결과 – 특허 출원 수



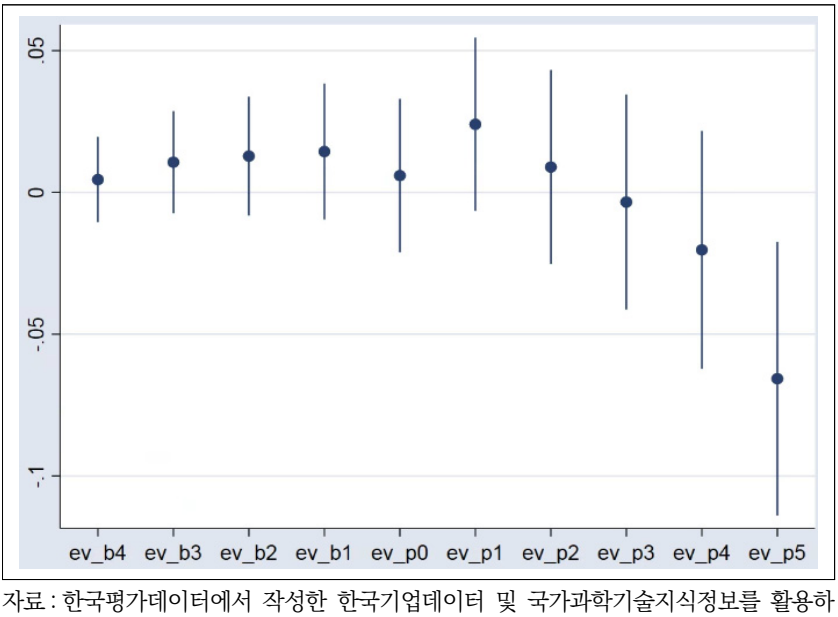
자료 : 한국평가데이터에서 작성한 한국기업데이터 및 국가과학기술지식정보를 활용하여 저자 작성.

고용은 어떻게 변하는지 살펴보는 것은 본 보고서에서 연구개발 활동 관련 사업이나 특허 제도에 대해 논의하고 정책적 시사점을 도출할 때 꼭 필요한 기초 자료라 할 수 있다.

[그림 4-2]에서는 특허 출원 수를 매출액으로 대체하여 사건연구 모형을 다시 추정하였다. 앞선 절에서와 마찬가지로 종속변수는 로그 변환하여 사용하였다. 추정 결과에 따르면, 사업 참여 이전에는 큰 변화가 없던 매출액이 국가연구개발사업의 지원을 받으면서 감소 추세에 접어든 것으로 나타났다. 다시 말해서 국가연구개발 지원사업 참여가 기업의 매출에 부정적인 영향을 미친 것으로 나타난 것이다.

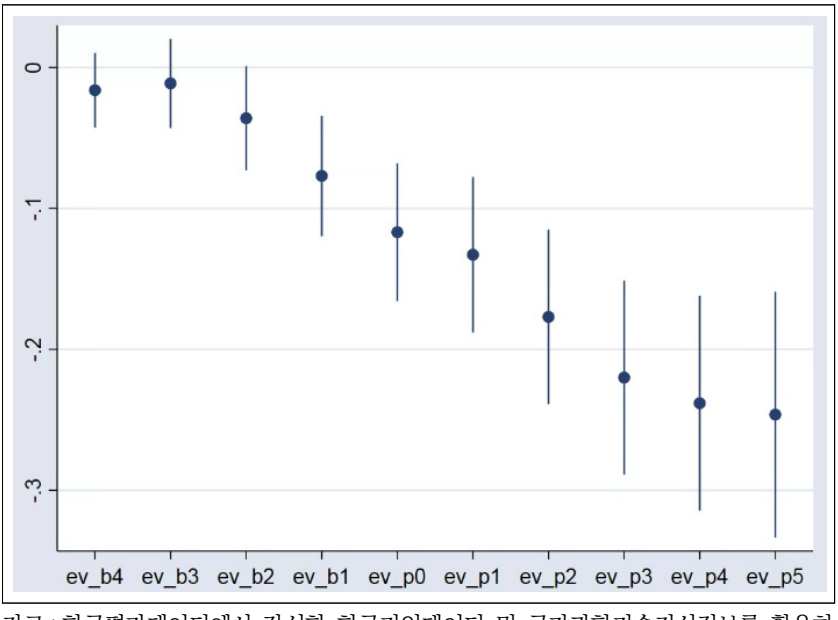
[그림 4-3]은 로그 변환한 당기순이익을 종속변수로 대체하여 추정한 결과이다. 당기순이익의 경우 사업에 참여하기 전부터 이미 감소하기 시작하며 이러한 감소 추세는 사업 참여 이후 3년째까지 이어지다가 4년째에 접어들면서 멈추었다. 따라서 국가연구개발 지원사업 참여가 기업의 당기순이익에도 부정적인 영향을 미친 것으로 보인다. 다만, 매출과 당기순이익상에서

[그림 4-2] 사건연구 모형 추정 결과 – 매출액



자료 : 한국평가데이터에서 작성한 한국기업데이터 및 국가과학기술지식정보를 활용하여 저자 작성.

[그림 4-3] 사건연구 모형 추정 결과 – 당기순이익

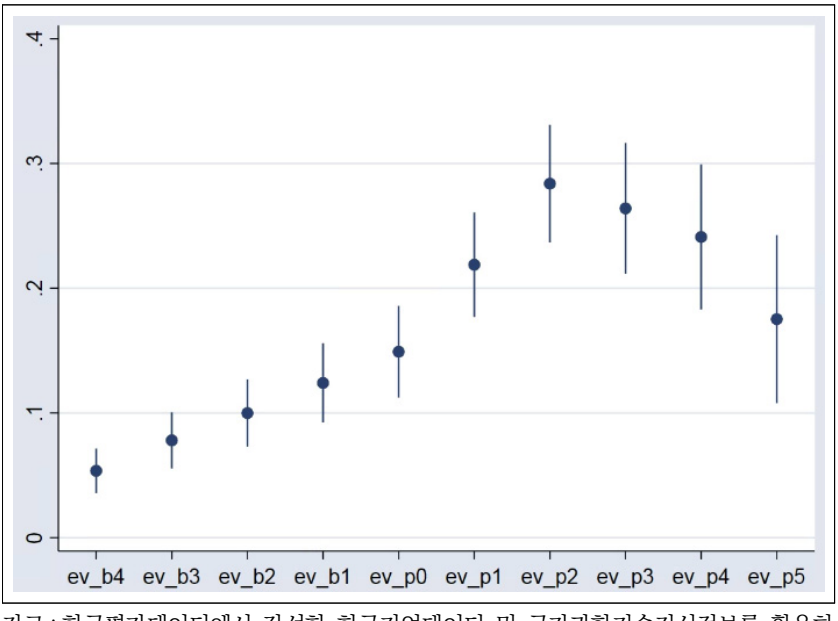


자료 : 한국평가데이터에서 작성한 한국기업데이터 및 국가과학기술지식정보를 활용하여 저자 작성.

사업 참여 이후 즉각적인 양의 효과를 얻지 못하더라도 국가연구개발 지원사업을 통한 기업에 대한 연구개발비 지원이 장기적인 연구개발 투자의 증가와 이로 인한 특허 출원 증가로 이어져 장기 생산성을 높인다면, 본 연구에서 분석한 지원사업 참여 이후의 시점보다 더 먼 미래의 매출과 이익으로 실현될 수도 있으므로, 추후 이에 대한 분석이 있기 전까지는 현시점에서의 분석 결과만을 가지고 국가연구개발 지원사업이 당초 목표한 소기의 성과를 거두지 못했다고 단정 지을 수는 없다.

[그림 4-4]는 로그 변환한 기업의 고용원 수를 종속변수로 놓고 사건연구 모형을 추정한 결과이다. 고용원 수는 지원사업 참여 이전에도 이미 증가 추세에 있었으며 사업에 참여한 순간 증가 추세가 소폭 가팔라진다. 그러나 사업 참여 시점으로부터 2년이 경과한 3년째까지만 이러한 고용 증가 추세가 지속되며 3년 후부터는 오히려 고용원 수가 감소하기 시작한다.

[그림 4-4] 사건연구 모형 추정 결과 – 고용원 수

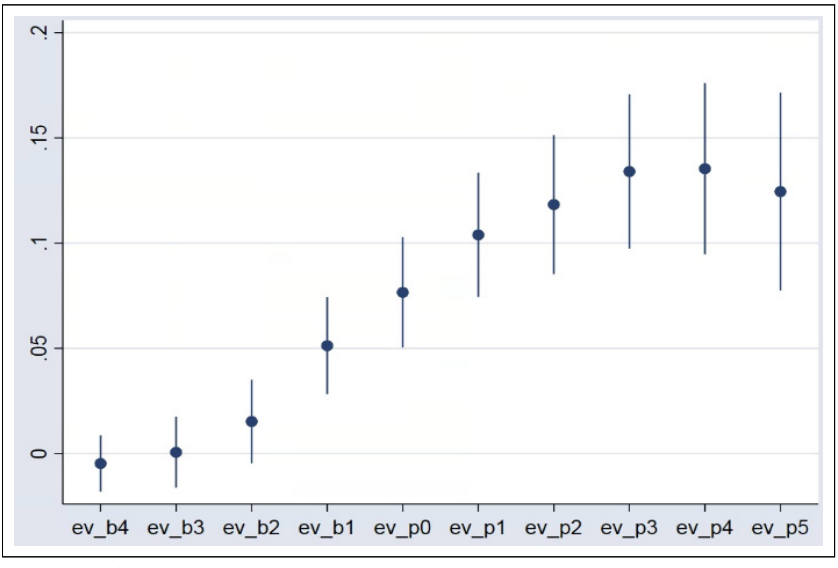


자료 : 한국평가데이터에서 작성한 한국기업데이터 및 국가과학기술지식정보를 활용하여 저자 작성.

로그 변환한 기업의 납입 자본금을 종속변수로 대체하여 추정한 결과를 [그림 4-5]에 그려 놓았다. 납입 자본금은 지원사업 참여 이전부터 증가하는 추세에 있었으며 이러한 추세는 정책 도입 이후에도 지속되는 것을 관찰할 수 있다. 그러나 이러한 증가 추세는 사업 참여 이후 3년이 경과한 4년째까지만 지속되며 4년 후부터는 증가세가 멈추고 정체되는 것으로 분석되었다. 다만, 사업 참여 이후 오랜 기간이 경과한 후에도 납입 자본금은 고용원 수와는 다르게 눈에 띄는 감소세가 나타나지는 않았다.

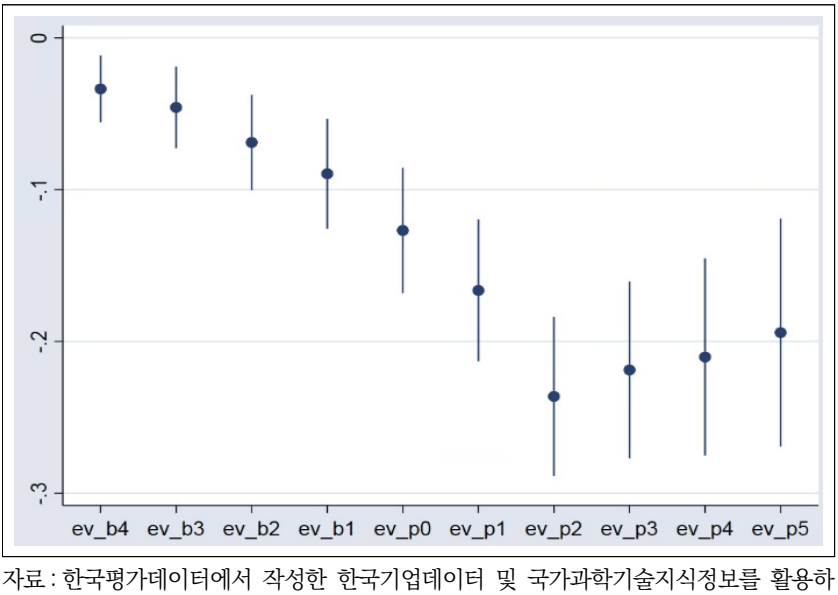
마지막으로 [그림 4-6]에서는 앞선 제3장과 유사하게 노동자 1인당 매출액을 노동생산성의 대리변수로 사용하여 사업의 생산성 효과를 추정하였다. 노동생산성은 사업 참여 이전에는 지속적으로 감소하고 있었으나 정부지원사업 참여 이후 상승세로 전환한 것이 눈에 띈다.

[그림 4-5] 사건연구 모형 추정 결과 – 납입 자본금



자료 : 한국평가데이터에서 작성한 한국기업데이터 및 국가과학기술지식정보를 활용하여 저자 작성.

[그림 4-6] 사건연구 모형 추정 결과 – 1인당 매출액



자료 : 한국평가데이터에서 작성한 한국기업데이터 및 국가과학기술지식정보를 활용하여 저자 작성.

제4절 소 결

본 장에서는 민간 분야의 연구개발 활동을 장려하고 촉진하기 위해 정부가 전개하는 대표적인 R&D 지원 사업인 국가연구개발 지원사업의 효과를 다양한 모형을 이용하여 추정하였다. 지원사업의 효과를 평가하기 위해 관찰한 종속변수는 연구개발 활동의 성과물인 특허 출원 수, 기업의 경영성과 지표인 당기순이익과 매출액, 생산성 지표인 1인당 매출액, 그리고 노동시장 변수인 고용원 수이다.

결과를 요약하면 다음과 같다. 첫째, 국가연구개발 지원사업에 참여 이후 기업의 특허 출원 수는 증가하지만 그 효과가 장기간 지속되지는 않으며, 더하여 이미 사업에 참여하기 전부터 특허 출원 수가 증가 추세에 있었던 것을 고려하면 이러한 증가는 사업의 효과라기보다는 연구개발 활동에 관심을 가지고 있거나 연구개발이 필요한 기업이 사업에 참여한 것으로 해석함이 타당하다 하겠다.

둘째, 다양한 기업의 경영성과를 결과 변수로 놓고 분석한 결과에 따르면, 고용원 수나 납입 자본금같이 중간 결과 변수의 성격을 지니는 생산 요소들은 사업 참여 초기까지는 증가하는 추세를 보이지만, 대체로 사업 참여 이후에 3년이 경과한 시점부터는 그 증가세가 둔화되거나 감소세로 접어들었다. 이를 통해 국가연구개발 지원사업 참여에 따른 연구개발 활동의 결과로 인한 생산성 증가가 비교적 단기에 관측되나 장기적으로는 기업의 자체적인 연구개발 활동이 중요하며 사업 효과가 영구적이거나 초장기에까지 지속되지는 않는다는 점을 확인할 수 있다. 그리고 이러한 사업 참여에 따른 생산성 증대 효과는 대략 사업 참여 후 3년 정도까지 지속되는 것으로 추정된다.

반면에 당기순이익이나 매출 같은 경영성과 지표는 사업에 참여한 이후에 오히려 악화되는 것으로 나타났다. 그러나 이것을 가지고 국가연구개발 사업 참여가 생산성이나 경영성과 지표의 개선에 무용한 것으로 해석하는 것은 바람직하지 않으며, 사업 참여에 따른 연구개발의 성과가 경영 지표에

반영되는 데에는 장기간이 필요할 수 있으므로 향후 보다 긴 시계열을 놓고 시행한 분석의 결과를 보고 사업의 효과를 평가해야 한다고 판단된다.

마지막으로 생산 요소 중 특히 고용원 수에 있어서는, 다른 생산 요소와는 달리 국가연구개발 지원사업에 참여한 사업체에서 사업 참여 전후로 모두 양의 효과가 관찰되어, 국가연구개발사업 참여는 사업체의 노동수요를 늘릴 가능성이 높으며, 아울러 종사자 수를 기준으로 기업의 규모를 키우는 효과도 있음이 확인된다. 이는 연구개발 활동을 확대함에 따라 기업에서 연구개발 인력에 대한 추가적인 수요가 발생함을 추정케 하며, 아울러 이러한 연구개발에 대한 투자는 일단 시작한 이후에 일정 시점 동안에는 지속될 가능성이 높음을 시사한다. 또한 연구개발을 위해 고용한 인력의 상당수는 고학력 고숙련 인력이므로 이들 다수가 정규직으로 채용되기에 고용 증대 효과가 상당 기간 지속될 가능성과 함께 양질의 일자리 창출에 있어서 연구개발 활동이 어느 정도의 효과가 있음을 암시한다.

제 5 장
자연어 처리를 활용한 특허 분석

제1절 서 론

1. 특허 분석의 중요성

수십 년 동안 특허는 기술혁신과 발명에 대한 권리를 보호할 수 있는 수단으로서 기업에서 중요한 무형 자산으로 여겨져 왔다. 특히 대부분의 새로운 혁신이 기술 개발에서 비롯되는 오늘날의 기업 환경에서 특허는 기술 진보와 다양성을 나타내는 자료로서 기술과 혁신의 확산에 있어서도 중요한 역할을 한다.

특허 분석(Patent Analysis)이라 함은 특허에 관련한 데이터를 체계적으로 검토, 평가, 해석하는 과정이다. 특허 분석은 지적재산권 관리와 기술혁신에 있어 매우 중요한 절차로 여겨진다. 이러한 분석은 연구개발(R&D), 사업 계획, 경쟁 분석에서 전략적 의사결정을 촉진하기 위해 수행되며 구체적인 수행 목적과 중요성에 대한 상세한 설명은 다음과 같다.

기본적으로 특허 분석은 발명가에게 지적재산권으로 부여된 법적 문서인 특허를 연구하는 과정이다. 이 분석은 간단한 검색에서 복잡한 법적 · 기술적 검토에 이르기까지 다양한 방식으로 이루어질 수 있으며, 아울러 특허의 유효성 평가, 잠재적 기술 침해 식별 및 기술 동향 평가 등까지 포괄할 수 있

다. 이러한 특허 분석의 주요 목적은 특정 기술 분야의 최신 기술 상태에 대한 통찰력을 얻는 것이다. 이는 기술 격차를 인식하고, 시장의 추세 및 시장의 주요 플레이어를 식별하는 데 도움이 된다. 또한 기업들은 연구개발의 방향을 결정하거나, 발명의 특허 가능성을 평가하고 특허의 침해를 방지하기 위한 전략을 수립하기 위해 특허 분석을 사용한다. 한편 경쟁자가 출원한 특허를 분석함으로써 기업은 경쟁 기업의 연구 초점, 개발 전략 및 미래 계획에 대한 정보를 얻을 수도 있다. 이러한 정보는 기업이 시장에서 경쟁 우위를 획득하거나 유지하는 데 필수적이다.

연구개발 분야에서 특허 분석은 기존 기술을 식별하고 노력의 중복을 피하는 데 요긴하게 활용될 수 있다. 이는 연구가 혁신적이며 기존 지식풀에 가치를 더하는 것을 보장한다. 또한 특허를 분석하는 것은 새로운 아이디어와 연구 접근 방식에 영감을 주는 기능도 가지고 있다.

특허 분석은 기술 예측 및 혁신을 추적하는 면에서도 중요한 역할을 한다. 특허 동향을 분석함으로써 기업들은 미래의 기술 발전을 예측하고 그에 따라 사업 전략을 조정할 수 있다. 또한 다양한 분야에서 혁신이 진행되는 상황을 추적함으로써 산업이 향하는 방향은 어디인지에 대한 통찰력도 제공한다. 한편 모든 산업에서 기술의 표준화와 세계화가 가속화되고 있는 만큼 특허 분석을 통해 세계 시장에서의 기술 개발에 대해 이해할 수 있으며, 역으로 특허 분석을 통해 국제 기술 표준에 대한 이해도를 제고함으로써 회사가 생산하는 제품이나 서비스를 이러한 국제 표준에 맞출 수 있다. 특허 분석은 거시경제적인 측면에서도 경제적·정책적 함의를 가진다. 특허 분석의 결과에 따라 혁신, 기술 이전 및 지적재산권에 대한 정부정책이 바뀔 수도 있고, 경제적으로는 국가 및 지역의 기술 경쟁력을 평가하고 제고하는 데 기여할 수도 있다.

법률적인 관점에서 특허 분석은 소송이나 라이선스 협상에 도움이 된다. 특히나 특허의 범위와 주장을 이해하는 것은 기업을 불필요한 침해 소송으로부터 보호할 수 있다. 상업적으로는 잠재적 라이선스 기회와 파트너십을 식별하는 데 도움을 줌으로써 새로운 수익원을 창출할 수 있다.

결론적으로, 특허 분석은 비즈니스, 법률, 기술 및 정책의 영역에서 광범위한 함의를 가진 다면적인 과정이라고 할 수 있다. 기업이 전략적 의사 결

정을 내리는 데 중요한 지침을 제공하고, 경제적으로 혁신을 촉진하며, 다양한 경제 주체들이 복잡한 지적재산권 환경을 탐색하고 이해하는 데 있어 그 중요성을 강조하지 않을 수 없다. 기술 개발의 속도가 갈수록 빨라짐에 따라 산업과 경제의 미래를 예측하고 올바른 모습을 형성하는 데 특허 분석의 역할은 점점 더 중요해질 것이다. 본 장에서는 자연어에 기반하여 특허를 분석함으로써 향후 특허와 관련된 연구에서 자연어 분석의 활용 가능성을 탐색하고, 아울러 자연어 기반 특허 분석이 특허와 연구개발 관련 제도와 정책을 개선하는 데 어떠한 기여를 할 수 있는지도 알아볼 것이다.

2. 특허의 신규성

어떠한 발명이 특허로 인정받기 위해서는 해당 특허가 산업상 이용가능성과 신규성 및 진보성을 갖추었다고 인정되어야 한다. 이 중 산업상 이용가능성을 제외한 특허의 핵심적인 가치인 나머지 두 가지 특성(신규성 및 진보성)에 대해 자세히 살펴보면 다음과 같다.

특허의 신규성이라 함은 해당 발명이 출원 전에 세상에 알려진 기술이 아니라는 것을 의미하는데, 「특허법」 제29조 제1항에 규정된 신규성에 대한 정의는 다음과 같다.

① 산업상 이용할 수 있는 발명으로서 다음 각호의 어느 하나에 해당하는 것을 제외하고는 그 발명에 대하여 특허를 받을 수 있다.
1. 특허 출원 전에 국내 또는 국외에서 공지되었거나 공연히 실시된 발명
2. 특허출원 전에 국내 또는 국외에서 반포된 간행물에 게재되었거나 전기통신회선을 통하여 공중이 이용할 수 있는 발명

특허의 진보성이란 출원 전에 그 발명이 속하는 기술 분야에서 통상의 지식을 가진 사람이 공지 기술로부터 용이하게 창작할 수 없는 발명을 의미하는데, 일반적으로 복수의 선행 기술의 여러 부분을 짜깁기한 형식의 발명에 대해서는 진보성이 결여된 것으로 판단하여 특허를 인정하지 않는다. 「특허법」 제29조 제2항에 규정된 진보성에 대한 정의는 다음과 같다.

② 특허 출원 전에 그 발명이 속하는 기술 분야에서 통상의 지식을 가진 사람이 제1항 각호의 어느 하나에 해당하는 발명에 의하여 쉽게 발명할 수 있으면 그 발명에 대해서는 제1항에도 불구하고 특허를 받을 수 없다.

본 연구에서 특허의 신규성이라는 용어는 앞서 「특허법」에서 언급된 신규성과 진보성을 아우르는 포괄적인 개념으로 사용할 것이다. 따라서 특허의 신규성을 판단하는 자연어 분석은 특허의 대상이 되는 기술이 새로운 특허로 인정받을 만한 새로운 기술인지 아닌지를 구분해 내는 기술이라 할 수 있다. 현재 특허의 신규성에 대한 판단은 전적으로 특허 심사관의 수작업에 의해 이루어지며 이로 인해 많은 시간과 비용이 소요되고 있다.

특허 신규성 예측은 여러 분야에서 활용될 수 있는 중요한 도구로, 특허 심사부터 기술개발, 투자 분석, 경쟁 분석, 법적 분쟁 해결에 이르기까지 다양한 영역에서 그 가치를 발휘할 수 있다. 예를 들어, 특허 심사 과정에서 신규성 예측 도구는 심사관들이 신속하고 정확하게 특허의 독창성을 평가하는 데 도움을 줄 수 있으며, 궁극적으로는 심사 과정을 효율화하고, 특허의 질을 향상시키는 데 기여한다.

기술개발 분야에서 연구개발 부서는 신규성 예측을 활용하여 자사의 기술이나 제품이 시장에서 어떤 차별화된 위치를 차지할 수 있는지 특허 출원 이전에 사전적으로 평가할 수 있다. 더 나아가 이러한 분석은 연구개발 전략과 특허 전략의 수립에 중요한 정보를 제공할 수 있다. 투자자나 기술 이전 전문가들은 특허의 신규성을 보다 객관적인 도구로 평가함으로써 해당 기술의 시장 잠재력과 투자 가치를 분석하는 데 도움을 얻을 수 있다.

보편적으로 신규성이 높은 특허는 더 큰 상업적 가치를 가질 확률이 높다는 점에서 특허나 기술에 대한 신규성 분석은 투자와 관련된 의사 결정에 중요한 영향을 미칠 수 있다. 경쟁사의 특허 포트폴리오를 분석할 때 신규성 예측은 경쟁사의 혁신적인 기술 발전의 방향을 예측하고, 자사의 경쟁 전략을 조정하는 데 도움이 된다. 이는 시장 내에서 경쟁상의 우위를 확보하는 데 중요한 역할을 할 수 있다. 또한 법적 분쟁이나 특허 침해 분석에서도 특허 침해 소송에서 특허의 신규성을 입증함으로써 침해 주장을 강화하거나 방어 전략

을 개발하는 데 사용됨으로써 신규성 예측은 중요한 역할을 할 수 있다.

마지막으로, 학계에서는 「특허법」, 지적재산권, 기술혁신에 대한 이해를 높이기 위해 특허 신규성 예측 연구를 활용할 수 있다.

3. 연구의 구성

본 장은 다음과 같이 구성된다. 우선 제2절에서는 특허에 대해서 자연어 처리를 활용하여 분석한 선행 연구들과 연구들의 방법론을 정리해 보고자 한다. 특허 정보 데이터는 자료의 크기 및 특수성으로 인해 정보의 접근 및 이해가 어렵다는 한계가 존재한다. 따라서 제3절에서는 특허 정보 데이터의 구성에 대해서 상세하게 설명하여 이후 분석 결과를 해석하는 데 있어서 독자의 이해를 도울 것이다. 이러한 작업에는 향후 본 연구에서 중요하게 활용될 선행기술조사의 의미와 선행기술 인용의 의미도 포함한다. 제4절에서는 실제 수집한 특허의 초록을 이용하여 특허의 신규성에 대해 자연어 분석을 실시한 결과를 보여줄 것이다. 동일한 자료를 이용하여 제5절에서는 피인용 횟수별로 특허를 나누어 실제 피인용 횟수의 차이에 따라 자연어 분석상에서 차이가 나타나는지를 살펴보고 궁극적으로는 피인용 횟수가 특허의 가치를 잘 판단하는 도구가 될 수 있는지 살펴보고자 한다. 제6절에서는 동일한 방법으로 2001년 이후 기계 · 설비 산업 분야에서 특허 등록량이 많은 기업 중 고용 변화율이 높은 기업들을 추려서 고용 변화와 특허의 자연어 상에서의 특성 간의 관계를 분석할 것이다. 마지막으로 제7절에서는 본 장의 결과와 향후의 활용 가능성을 정리한다.

제2절 선행 연구

1. 자연어 처리 기법을 이용한 특허 분류

특허 분석과 관련된 여러 연구 분야 중 기계학습 방법에 기반한 자연어

처리 기법을 많이 사용한 분야는 특허 분류와 관련된 연구이다. 특허 분류 작업은 특허 문서를 국제 특허 분류(International Patent Classification : IPC) 혹은 선진 특허 분류(Cooperative Patent Classification : CPC) 체계에 따라 분류코드를 할당해야 하는 특허 분석 작업이다. 특허를 체계적으로 분류하는 것은 특허를 효율적으로 관리하고 탐색을 원활하게 하는 데 있어 필수적이다.

오랫동안 특허 문서는 수동으로 분석되고, 그 후에 신청자와 특허 담당자에 의해 분류코드가 할당되어 왔다. 이러한 수동 라벨링(labeling) 작업은 전문적인 도메인 지식(domain knowledge)이 필요하며 시간이 많이 소요될 뿐만 아니라 인간에 의한 실수 등 오류 가능성도 내포하고 있다. 특허 분류 작업의 특수성에 더해 시간이 지나면서 검토해야 할 특허의 수가 급격히 증가함에 따라, 자연어 처리를 활용하여 특허를 자동으로 분류할 수 있는 방법을 찾기 위한 연구가 수행되었다.

방법론적으로 이 연구에는 나이브 베이즈(Naïve Bayes), k-최근접 이웃(k-Nearest Neighborm : kNN), 서포트 벡터 머신(Support Vector Machine : SVM), 의사결정 나무(Decision Tree)와 같은 전통적인 기계학습 방법들이 많이 사용되었고, 최근 들어 심층학습 기술을 적용하려는 시도도 꾸준히 이루어지고 있다.

특허 분류에 대한 연구는 분류코드 체계가 계층적이기 때문에, 설정에서 다양한 변형이 존재한다. 예를 들어, 어떠한 연구는 기계학습을 활용하여 특허 분류 체계의 최상위 클래스에 해당하는 코드만 정확하게 예측하도록 한다. 최상위 클래스를 예측하는 것은 분류 프로그램 중에서 가장 간단한 버전이며 정확성은 높지만 가장 유용성이 적다. 보다 유용하고 어려운 설정은 특정 수준까지 이하의 하위 클래스 분류까지 정확하게 예측하는 것이다. 이러한 설정은 CLEF-IP 2010 등과 같은 대규모 경진 대회에서도 사용되었다.

한편 특허 분류 프로그램은 이미 완성되어 사용 중인 특허 분류 체계를 평가하고 개선하는 데에도 활용할 수 있다. 어떠한 특허가 하나의 하위 분류가 아닌 여러 가지의 하위 분류에 할당될 수 있는데, 따라서 분류 프로그램을 이용하여 현재 사용하고 있는 분류 체계가 얼마나 체계적이며 현재의 기술 발전에 비추어 체계적인지 평가할 수 있으며, 현재의 분류 체계를 정리

하고 새로운 하위 분류를 도입하는 데에도 사용할 수 있다. 예를 들어 설명하자면, 특허 분류 체계를 평가하는 가장 직관적인 방법은 특허 분류 체계 하에서 가장 높은 확률로 예측된 하위 분류가 특허에 실제로 할당된 분류에 속해 있으면 합리적인 체계로 간주하는 방법으로 평가하는 것이다. 또 다른 평가 방법은 예측 순위에서 상위 세 가지 하위 분류에 대해서 실제로 할당된 분류 중 몇 가지와 매칭되는지를 평가할 수도 있다. 다만 단순히 '어떠한 시스템의 성능 지표(예를 들면, 정확도)가 몇 %이다'라고 하는 결과만을 가지고 특정 시스템이 다른 시스템보다 우월하다고 말할 수는 없다.

가. 초기 연구

특허 분류에 관한 초기 연구 중 하나인 Larkey(1999)는 각 특허를 특정 키워드의 발생 빈도 형태로 표현하고 이를 각 특허의 특징(feature, 기계학습에서 예측하고자 하는 변수에 대한 독립 변수에 해당하는 개념으로 활용하는 용어)으로 이용하여 예측을 위해 k-최근접 이웃을 활용한 분류기를 구축하여 사용했다.

Fall et al.(2004)은 독일어 특허 분류 작업에서 전문가 시스템을 훈련하기 위해 나이브 베이즈, k-최근접 이웃, SVM과 같은 다양한 기계학습 알고리즘을 사용했다. 그들은 문서를 표현하기 위해 두 가지 대안적 색인 접근법을 사용했다.

Li and Shawe-Taylor(2007)는 특허 표현을 위해 TF-IDF(Term Frequency-Inverse Document Frequency) 벡터 표현을 사용하고, 예측을 위해 커널 정준상관분석(Kernel Canonical Correlation Analysis : KCCA)과 SVM을 적용했다. Khattak and Heyer(2011)에서는 특허 분류를 위해 저빈도 용어의 사용과 그 특성에 중점을 두고 분류 시스템을 완성하였다.

나. 주제 분석 기법의 사용

주제 모델링(Topic Modelling)은 키워드의 확률 분포를 기반으로 대규모 데이터베이스에서 주제를 추출하는 방법이다(Blei, Ng and Jordan, 2003).

주제 모델링의 대표적인 방법은 잠재 디리클레 할당(Latent Dirichlet Allocation : LDA)으로, 코퍼스(Corpus)의 생성 확률을 측정하여 주제를 식별하는 모델이다. LDA 모델은 어떠한 문서에서 의미 있는 주제를 식별하는 데 매우 효과적이기 때문에 LDA를 이용한 주제 모델링은 많은 연구에서 널리 사용되었다. Moro, Cortez and Rita(2017)는 관련 학술 기사를 사용하여 은행업계와 관련된 특허 주제를 식별하는 연구를 수행했다.

주제 분석은 기술 동향을 분석하기 위한 주요 방법으로서도 활용되고 있다. Chen et al.(2015)은 특허의 주장을 사용하여 기술 발전과 관련된 주제 변화를 식별하기 위해 LDA를 사용했다. Chen et al.(2017)의 연구에서도 기술 동향을 식별하기 위한 주제 기반 기술예측에 동일한 LDA 접근 방식을 제안했다. 이 연구에서는 시간적 동향 패턴과 의미 있는 주제가 식별되었으며 주제 기반 지식과 동향 예측을 제공했다. Kim, Park and Yoon(2016)은 LDA를 활용하여 기술 모니터링 목적의 특허 지도 개발 방법을 다루었다. 그들은 다양한 기간을 대상으로 기술 주제를 추출하기 위해 LDA를 사용했다.

Suominen, Toivanen and Seppänen(2017)은 주제 모델링을 포함한 비감독 학습 사용 방법을 제안했다. 해당 논문에서는 LDA로 추출된 주제를 각 시기별로 시각화하여 미래를 예측하는 데 사용하였다. Ranaei and Suominen(2017)은 차량 관련 특허의 신흥 패턴을 식별하기 위해 LDA와 동적 주제 모델링을 사용했다.

특허 분류와 관련하여, Venugopalan and Rai(2015)는 특허 분류 및 패턴 식별을 위해 주제 모델링을 사용했다. 그들은 특허의 자동 식별 및 분류를 위한 자연어 처리 기반 계층적 기술을 사용했다. 해당 연구에서는 첫째로 관련 특허를 식별한 후, 그다음 과정으로 언어 패턴에 기반한 관련 특허에 대한 주제 모델링을 포함하는 두 단계 프레임워크를 사용했다.

다. 전통적인 기계학습 기법의 사용

전통적인 기계학습 중 특허 분류에 가장 많이 사용된 방법은 서포트 벡터 머신(Support Vector Machine : SVM)이었다. SVM은 분석 대상들의 다양한 클래스를 분류하기 위한 최적의 초평면을 식별하는 데 사용되는 대표적인

기계학습 기술이다(Vapnik, 2013). 특히나 분류, 회귀 및 특징 선택(feature selection)에 있어 SVM의 성능은 뛰어난 것으로 알려져 있다.

SVM은 특허 분류뿐만 아니라 다른 분류 작업에서도 가장 인기 있는 분류 기술 중 하나이며, 상대적으로 우수한 일반화 성능을 보였다(Kang et al., 2015). 이 때문에 텍스트 분류, 이미지 분할 및 예측, 의료 관련 분류와 예측을 포함한 많은 분류 문제에 사용되었다. 최근에는 SVM을 단순한 데이터 분류에만 활용하는 데 그치지 않고 기술예측 분야에서도 적극적으로 사용하고 있다. Yoon and Magee(2018)는 향후 유망한 기술 분야를 선택하기 위한 링크 예측에 SVM을 사용했다. 이 연구에서는 두 특허의 키워드에서의 빈도 차이와 키워드 벡터라는 독립변수를 기반으로 연결 유무가 식별된다.

Kyebambe et al.(2017)은 신흥 기술을 예측하는 데 SVM을 포함한 감독학습 기술을 사용했다. 신흥 기술과 숙성된 기술을 구별하기 위해 청구 수, 인용 수, 기술 주기 시간, 특허 클래스, 인용 기술 유사성 지수, 인용 특허 양도인 유사성 지수를 비롯한 여러 특징을 사용했다. Kong et al.(2017)은 데이터 마이닝 기술과 전문가의 질적 지식을 결합한 후, 혁신에 유용한 고품질 특허를 식별하기 위해 SVM 기반 분류기를 사용했다. Lee et al.(2018)도 신흥 기술의 조기 식별을 위해 잠재적 영향을 출력 지표로, 기술적 특성을 입력 지표로 사용하는 SVM을 사용한 바 있다.

라. 심층학습 기법의 사용

특허 분류 작업에 자연어 처리 기반 심층학습 기법을 처음 도입한 것은 Grawe et al.(2017)이었다. Grawe et al.(2017)은 텍스트 기반 특허 분류를 위해 심층학습과 단어 임베딩 방법을 도입했다. Hepburn(2018)은 특허 분류를 위해 변환기 모델을 사용하는 아이디어를 제안했다. 서포트 벡터 머신(SVM)과 범용 언어 모델인 ULMFiT를 사용하여, IPC(섹션 수준, 8개 라벨) 예측에 대해 78.4%의 F1 점수를 달성했다. Li et al.(2018)은 특허 분류를 위해 특허 제목과 초록의 단어 벡터 임베딩을 기반으로 한 합성곱 신경망(CNN)을 사용한 심층학습 알고리즘인 DeepPatent를 제안했다. 해당 연구의 결과에서 F1 점수는 약 43%였다. 이 연구에서는 DeepPatent가 자동 특

징 추출로 73.88%의 분류 정확도를 달성했으며, 같은 정보를 훈련에 사용한 기존의 모든 알고리즘보다 더 나은 성능을 보였음을 보여주었다. 논문에서는 IPC 하위 클래스 수준의 637개 범주에 대한 267만 9,443개의 미국 특허 문서 중 데이터를 정리한 후 200만 147개 특허 문서에 대해서 USPTO-2M 데이터에서 평가를 수행했다(Li et al., 2018).

Lee and Hsiang(2020)은 BERT 사전 훈련된 모델을 기반으로 한 트랜스포머 모델인 PatentBERT를 특허 분류를 위해 제안했다. 해당 연구에서는 BERT 모델을 특허 분류를 위해 미세조정하는 데 중점을 두고, CPC 하위 클래스 수준의 200만 개 이상의 특허로 구성된 대규모 데이터인 USPTO를 사용했다. 연구에서는 제목과 초록을 제외한 특허의 클레임 정보만을 사용하여 다른 분류 방식들과 비슷한 F1 스코어를 보임으로써, 클레임 정보만으로도 특허 분류 작업을 충분히 수행할 수 있음을 입증하였다. 여기서 제목과 초록까지 포함하여 분석한 F1 Top5 결과는 44.75%였다. 이 접근 방식에서 저자들은 적절한 정확도를 가진 모델을 제안했지만, 특허 간의 유사성 측정 방법이나 이를 수행하기 위한 워크플로를 담거나 제안하지는 않았다는 한계가 존재한다(Hain et al., 2021).

2. 기술예측 또는 동향 분석

기술예측(Technology Forecasting : TF)이란 특허 정보에 포함된 기술을 분석하여 향후 유용할 것이라고 예상되는 기술의 특징을 예측하는 작업이다. 기술예측은 경영진이 기업의 전략을 수립하거나, 연구개발 활동의 관리, 제품개발, 새로운 공정 기술에 대한 투자, 생산 및 마케팅, 새로운 기술의 구매 등에 있어 보다 나은 의사 결정을 내리는 데 도움을 주기 때문에, 이 분야에 자연어 처리를 도입할 경우 기업 경영에서 정확성과 신속성, 그리고 적시성을 확보할 수 있다.

전통적으로 기술예측에 활용되어 온 도구 중 하나는 S-곡선으로, 기술이 나타난 후 시간이 흐름에 따라 축적된 특허의 수를 가지고 현 시점에서 해당 기술의 단계를 보여주는 곡선 유형이다(Altuntas, Dereli and Kusiak, 2015). S-곡선은 시작, 성장 및 포화의 세 단계로 구성되어 있다. 새로운 기

술 분야에서는 왕성한 혁신 활동으로 인해 특허의 수가 빠르게 성장하지만, 이후 그 속도가 서서히 줄어들었다가 기술이 성숙하고 포화 상태에 다다르면 더 이상 많은 신규 특허가 출원되거나 등록되지 않는 단계를 따른다. 해당 분야에서 특허가 출원되는 개수를 기간별로 분석하면 S-곡선을 작성할 수 있으므로 이를 사용하여 연구자들은 특정 기술이 어느 시점에 있는지 판단하였다. 일반적으로 시작 및 포화 단계에서의 투자는 위험성 및 수익성 등 여러 이유로 권장되지 않으므로, 투자는 성장 단계 동안에 이루어지는 것이 가장 합리적이다.

기술 예측 분석을 위한 다른 방법으로는 워드 클라우드를 활용한 방법이 있다(Bamakan et al., 2021). 워드 클라우드는 텍스트 데이터에서 사용되는 고빈도 용어를 빠른 시각 안에 표시하여 기술예측에서 널리 사용된다(Yang and Hwang, 2020). 해당 연구에서 사용된 모형은 시점별로 특정 분야의 특허 문서에 포함된 키워드들을 추출하여 빈도가 높은 단어를 시각적으로 보여줌으로써 기술의 변화 양상을 파악할 수 있게 해준다.

특정 분야와 관련된 특허 정보를 수집하여 S-곡선 또는 워드 클라우드 기법을 활용하기 위해서는 유사한 특허들을 수집할 수 있는 자연어 처리 기법을 필요로 한다. 특허 정보 간의 유사성은 특허의 정보에 포함된 텍스트 정보의 유사성으로 측정을 하게 되는데, 텍스트 사이에 유사성을 측정하는 데는 다양한 방법이 있으며, 선택된 방법에 따라 각각 다른 결과를 나타내게 된다.

Li et al.(2021)의 연구는 의료 분야에서 주제-동작-대상(SAO)의 구분체계를 도입하여 유사한 기술을 측정하는 방법을 제시하였으며, 알고리즘을 실제로 구현해 보기 위해 알츠하이머 질환에 대한 경험적 연구를 수행했다. 연구에서 사용한 방법은 관련된 특허를 수집하는 데 있어 키워드와 국제 특허분류에 의한 전통적인 방법보다 더 신뢰할 수 있는 결과를 제시했다. Tan, Zhao and Zhang(2022)의 연구는 기술 문서의 표현을 학습하기 위한 결합된 텍스트 쌍 임베딩(Combined Text Pair Embeddings : CTPE) 모델을 제안했다. 이 모델은 텍스트 쌍의 결합 여부를 판단하는 것을 목적으로 훈련된다. 해당 연구에서는 CTPE 모델의 성능을 평가하기 위해 word2vec과 같은 임베딩 모델, 단어 수준의 표현 방법(LSA, WMD 및 기타), 그리고 문장 수준 표현

방법, 이렇게 세 가지 유형의 비감독 분산 문서 표현 방법과 비교하였다.

Kim et al.(2018)은 기업 기술의 쇠퇴를 예측하기 위해 게임 이론을 적용한 동적 프레임워크를 제안했다. 이 연구에서는 HD-DVD와 블루레이 간의 표준 전쟁 사례가 분석되었으며, 기업의 기술 관계에서의 변화를 관찰하여 기업에 대한 소비자들의 선호도 하락과 이로 인한 시장 균형의 변화를 예측할 수 있었다. Xu, Mu and Chen(2020)은 짧은 텍스트를 분석하기 위한 다중시각 유사성 측정 프레임워크를 제안했다. 다중시각 유사성 측정 프레임워크는 의미론적 유사성이나 구문 기반의 측정을 포함한 여러 종류의 기존 유사성 측정 방법을 쉽게 통합한 방법이다. 이 프레임워크의 또 다른 중요한 구성 요소는 기계학습 기반 통합 정책을 사용하는 유사성 측정 통합 모듈이다. 실제 실험은 대규모 기업 IT 인프라에서의 실제 티켓 데이터 세트를 사용하였다.

3. 특허 신규성 예측

신규성은 문자 그대로 기술혁신 분야에서의 창의성 또는 새로움을 나타낸다(Plantec et al., 2021; Small, Boyack, and Klavans, 2014). 신규성 개념은 재조합(recombination)과 정제(refinement)로 나눌 수 있다(Strumsky and Lobo, 2015). 이전 연구들은 특허 자료에 포함된 텍스트나 메타 정보를 활용하여 신규성을 평가하는 데 도움이 되는 프레임워크를 제안했었다. 방법론적으로는 특허 기반의 문헌학적 신규성 지표, 이상치 탐지, 심층학습 기반 신규성 예측 등이 주로 사용되었다.

문헌학적 신규성 지표를 활용하는 연구에서는 각 기술 분야의 분류코드를 활용하여 신규성의 정도를 측정하였다. 특허 분류 체계에서는 새로운 하위 분류 간의 조합이 발견되면 신규성이 있는 것으로 판단하곤 한다(Plantec et al., 2021; Verhoeven et al., 2016). 이상치 탐지를 활용하여 특허의 신규성을 판단하는 방법은 텍스트 정보가 포함된 특허 지도를 생성하여 특허를 검토하는 접근 방법 중 하나이다(Wang and Chen, 2019; Zanella et al., 2021). 이렇게 생성된 특허 지도에서 혼잡한 지역, 즉 이미 많은 기술들이 출현하여 밀집해 있는 지역에서 멀리 떨어진 특허는 연구개발에서의

새로운 영역으로 식별되어 신규성이 높은 것으로 인식할 수 있다(Lee et al., 2015).

다른 접근 방법들은 주로 특허의 텍스트 정보를 사용하여 이미 등록된 특허 정보와의 유사성을 계산하는데, 이는 신규성 분석을 위한 가장 일반적인 접근 방법이다. 최근에는 이를 위해 심층학습 모델이 주로 사용되며, 자연어 처리 모델과 라벨이 지정된 대량 데이터를 사용하여 신규성을 분류하려는 연구가 주를 이루고 있다(Risch, Alder, Hewel and Krestel, 2020). 이 중 몇몇 연구들은 출원하고자 하는 특허와 이미 등록된 특허라는 두 가지 특허로 이루어진 특허 쌍을 비교하여 출원하는 특허가 이미 등록된 특허에 손해를 입힐 가능성은 없는지를 판단하는 감독학습(Supervised Learning) 모델을 제안했다(Freunek and Bodmer, 2021a; 2021b). 또 다른 연구에서는 인코더-디코더(Encoder-Decoder) 접근 방식이 사용되었는데, 라벨이 없는 특허 텍스트를 사용하여 두 특허 간의 기술적 상관관계를 추정했다(Chikkamath et al., 2020; Risch, Alder, Hewel and Krestel, 2020).

상술한 연구들은 자연어 처리 기법과 심층학습에 기반을 둔 접근 방법을 이용하여 특허의 신규성을 어떻게 분석할 수 있는지 보여주지만, 현 시점에서 이러한 방법론의 수준은 초기 단계에 머물러 있는 것으로 보인다. 현실적으로 기술 문서(technical documents)를 다룰 때에는 여러 가지 한계점이 있다. 특허와 같은 텍스트 기반의 기술 데이터는 다른 일반 문서들과는 달리 특수한 성격을 지닌 문서이다(Jang et al., 2021). 인공지능 모델이 특허에서의 특세트 입력을 인코딩할 때, 복잡한 기술 용어뿐만 아니라 청구항의 복잡한 구조도 포함해서 처리를 해야 하는 난제(難題)가 있는데, 이것은 앞으로 자연어 처리 기법을 발전시켜 해결해야 한다.

4. 특허 분석에 활용된 자연어 기법

본 소절에서는 앞서 이루어진 특허 분석에서 사용하는 자연어 처리를 이용한 연구를 접근 방법에 따라 분류하면서 세부 기술들에 대해서 간략하게 소개함으로써 이후 본 장에서 수행할 분석에 대한 기초 지식을 제공하고자 한다. 특허 및 신규성 분석 시에 고려한 요소 및 분석 방법론을 접근 방법별

로 구분하면 〈표 5-1〉과 같다.

특허는 기본적으로 자연어로 구성되어 있다. 특허의 본문뿐만 아니라 초록도 자연어로 이루어져 있으며, 따라서 특허가 가지는 가치 및 신규성 역시 모두 출원된 특허를 표시하는 자연어에서 찾을 수 있어야 한다. 그러므로 텍스트 데이터에서 정보를 추출하고 처리하는 데 필수적인 자연어 처리는 대량의 복잡한 특허 문서를 다루는 데 도움이 된다. 특허 분석에서 사용하는

〈표 5-1〉 자연어 처리를 활용한 특허 연구 방법론 분류

접근 방법	신규성 분석 시 고려 요소	분석 대상 및 방법
서지 계량화에 기반한 신규성 지표	- 재조합 및 지식 기원 - 기술 영향 분석 - 기술적 독창성 - 비즈니스에서의 지식 검색 방식	- 국제 특허 분류 코드 및 인용 정보 - 특허 가족 내 분류 코드 쌍
이상치 탐지를 통한 새로운 특허 탐색	- 신규성의 정도(또는 이상치로서의 특성) - 특허 지도를 활용한 기술 기회 발견 - 예상된 주요 행동에서 벗어난 이상한 패턴 - 특허 지도 및 사용자 유틸리티 지도를 활용한 연구개발 계획 - 원본이거나 이상한 상태(평균 의미적 유사도에서의 편차) 기술 동향 분석	- 기술 키워드 형태론을 사용한 밀도 기반 이상 탐지 방법 - 키워드 벡터를 사용한 각도 기반 이상치 탐지 방법 - 코사인 기반 및 밀도 기반 이상치 분석
텍스트 분석 기반 특허 유사도 지수	- 재조합 및 독특성 - 특허 평가 - 연속체의 양극으로서의 재조합 및 개척적 신규성 - 특허 신규성 검토	- 워드 임베딩 및 인용 기반 특허 유사성 - SAO(Subject-Action-Object) 기반 특허 유사성
자연어 처리와 딥러닝을 사용한 신규성 예측 모델 (지도 학습 방법)	- 신규성 검색 - 특허 신규성 검토 - 기존 기술 판별자 - 발명의 독특성	- BERT 기반 임베딩을 사용한 두 특허 간의 신규성 관련성 예측 - word2vec 단어 임베딩과 딥러닝 모델을 사용한 신규성 분류 - 단어 임베딩과 지도 학습 모델을 사용한 신규성 분류

자료 : 저자 작성.

주요 자연어 처리 기법을 정리하자면 다음과 같다.

가. 텍스트 마이닝(Text Mining)

텍스트 마이닝은 텍스트에서 가치 있는 정보를 추출하는 것을 지칭한다. 텍스트 마이닝은 특허 데이터베이스에서 일정한 추세나 패턴, 혹은 기술적 통찰을 발견하는 데 특히 중요하다. 특허 문서에 존재하는 주제를 식별하고 다른 주제와의 관련성을 파악할 수 있는 주제 모델링과 같은 기술이 여기에 속하며, 대량의 특허 데이터를 분류하고 요약하는 데 활용된다.

나. 감성 분석(Sentiment Analysis)

전통적으로 소비자 피드백이나 소셜 미디어 데이터와 더 관련이 있는 감정 분석도 특허 분석에서 널리 활용되고 있다. 감성 분석은 연구 논문이나 특허 출원의 어조(語調)와 방향을 평가하여 다양한 기술의 가치와 잠재적 영향을 분석할 수 있게 해준다. 또한 특정 기술에 대한 공개 문서나 경쟁사 분석에 있어서 사람들이나 경쟁 기업의 인식을 평가하는 데에도 활용 가능성이 있다.

다. 개체명 인식(Named Entity Recognition : NER)

개체명 인식은 특허 문서 내의 발명가, 조직, 기술 용어와 같은 핵심 개체를 식별하고 분류하는 데 필수적이다. 특허 분석에 있어 이 기술은 특허 소유권, 협업 및 기술 분야를 식별하는 데 주로 사용된다.

라. 의미 분석(Semantic Analysis)

의미 분석은 텍스트 내 단어의 의미와 맥락을 이해하기 위한 방법이다. 특허 분석에서는 복잡한 기술적 언어를 해석하고 관련 정보를 추출하는 것

을 포함한다. 의미 분석은 특허의 범위와 주장을 이해하기 위해 필수적인 방법이다.

마. 분류 및 클러스터링(Classification and Clustering)

이 기술은 특허를 관련 집단이나 층위로 분류하는 데 사용된다. 분류는 특허를 사전 정의된 분류 체계에 할당하는 작업이며 효율적인 특허 관리와 검색에 도움을 준다. 이 방법을 사용하면 특허를 기술 유형이나 응용 분야 또는 기타 관련 기준에 따라 분류할 수 있으며, 특정한 주제어나 단어에 대한 검색과 분석도 가능하다.

클러스터링은 앞서 설명한 분류와는 반대로 내용을 기반으로 유사한 문서를 유형화하여 특허 포트폴리오를 조직화하고 간소화하게끔 해준다. 또한 특허의 내용을 기반으로 유사한 특허를 묶어서 비교 분석을 보다 용이하게 하고 특정 기술 분야에서 특허를 유형화하여 식별 가능케 한다. 특허 간의 패턴이 명확하지 않거나 식별에 어려움이 있는 경우에도 패턴과 관계를 밝히는 데 클러스터링을 활용할 수 있다.

바. 정보 추출(Information Extraction)

정보 추출은 특허에 대한 정보에서 특허 번호, 날짜, 기술 사양과 같은 텍스트 내의 특정 정보나 특정 키워드를 추출해 내는 방법이다. 키워드 추출은 특허 내에서 중요한 단어나 구문을 식별하는 데에도 활용할 수 있으며, 정보 검색 및 주제 식별을 쉽게 할 수 있도록 도와준다. 특허와 관련된 데이터베이스를 자동으로 식별하여 채우고 특허 문서를 요약하는 것을 자동화하는 데 정보 추출 방법이 유용하게 활용될 수 있다.

사. 기계학습 및 심층학습 모델

최신 자연어 처리 기술은 신경망과 같은 기계학습 및 심층학습 모델을 활용한다. 이러한 모델들은 대량의 데이터 세트를 학습하고 그 결과 특허 언어

의 미묘함과 복잡성까지 다룰 수 있다. 특히 심층학습을 기반으로 한 최신의 초거대 언어 모형(Large Language Model : LLM)들은 기존에는 생각할 수 없었던 양의 파라미터를 사용하여 (일례로, ChatGPT의 경우 1,750억 개의 파라미터를 사용) 언어의 미묘함을 효과적으로 통제할 수 있다.

아. 자연어 생성(Natural Language Generation)

자연어 생성 방법을 활용하면 특허 자료에서 일관된 요약이나 설명을 자동으로 생성할 수 있다. 실용적으로는 특허의 요약을 빠르게 작성하고, 의사결정자를 위해 특허 포트폴리오에 대한 간략한 보고서를 생성하는 데 많이 사용되고 있다.

자. 단어 임베딩과 변환기 모델(Word Embedding and Transformer Models)

Word2Vec, BERT, GPT와 같은 기술은 자연어 처리 분야에 일대 혁신을 가져왔다. 특허 분석에서는 이러한 모델을 특허 유사성 분석과 같은 작업에 사용한다. 이 기법은 특허 문서 내에서 단어의 맥락적 관계를 이용하여 해당 특허와 관련된 문서나 혁신이 무엇인지를 식별하는 데 활용할 수 있다.

상술한 자연어 처리 기술들을 특허 분석에 활용하면 특허 데이터를 처리하는 과정을 간소화할 수 있을 뿐만 아니라 각종 의사결정이나 특허와 관련된 다양한 배경에 대한 이해를 높일 수 있다. 자연어 처리 기술이 지속적으로 발전함에 따라, 특허 분석에서 자연어 처리 기법의 활용은 더욱 정교하고 세밀한 영역까지 확대될 것이며 또한 필수적인 부분이 될 것으로 예상된다. 본 보고서에서는 특허 분석을 다양한 자연어 처리 기법을 이용하여 실시함으로써 그 가능성을 확인하고, 특히 한국어에 대해서도 자연어 처리 기법이 잘 활용될 수 있는지를 파악해 보고자 한다.

제3절 특허 데이터

1. 특허 문서에서 제공하는 세부 정보

특허 문서에는 다양한 종류의 정보가 포함되어 있어, 거의 모든 주요 사항을 확인할 수 있다. 이 정보는 다음과 같이 분류할 수 있다.

가. 개인 정보

발명자와 출원자의 성명 혹은 법인명이 명시되어 있으며, 특허를 대리인을 통해 신청한 경우 대리인의 정보도 담겨있다. 또한 등록된 특허 문서에서는 해당 기술을 검토한 심사관의 이름도 확인 가능하다.

- 출원자 및 발명자 : 출원자와 발명자는 대체로 동일하지만 간혹 다르다. 이러한 상황은 대학이나 연구 기관에서 일하는 연구자나 과학자들이 자신의 발명품을 대학 혹은 회사에 양도하는 경우 발생한다.
- 대리인 정보 : 특허 출원 과정은 복잡할 수 있기 때문에 많은 발명자들이 변리사를 포함한 특허 대리인의 도움을 받는다. 대리인은 출원 과정을 원활하게 진행하고 법적인 문제를 예방하는 데 도움을 준다.

나. 식별자

특허 문서에는 고유한 식별번호가 포함되어 있어, 해당 문서와 기술을 다른 특허의 문서나 기술과 구분할 수 있다. 여기에는 출원번호, 공개번호, 그리고 등록번호가 포함된다.

- 국가코드 : 일반적으로 식별번호에는 국가코드가 포함되어 있어, 해당 특허가 어느 국가에서 출원되었는지 알 수 있다.

• 종류코드 : 특허 문서의 종류(출원, 공개, 등록 등)를 나타내는 코드도 포함된 경우가 있다.

다. 날짜 정보

특허의 존속 기간은 출원일로부터 20년이며, 만료 이전까지 여러 중요한 시기들이 특허와 관련된 텍스트에 기입되어 있다. 여기에는 특허 출원일, 공개일, 공고일, 그리고 등록일이 있다. 공개일 이후에는 해당 특허가 공중에 공개되며 참고문헌으로 사용될 수 있다.

• 우선권 주장일 : 발명자가 특정 국가에서 먼저 특허를 출원한 후 등록된 특허에 대해서 다른 국가에서 출원하는 경우, 특허에 대한 우선권을 주장할 수 있다. 우선권 주장이 인정될 경우 해당 특허의 시효를 앞당길 수 있기 때문에 해당 국가 내에서 이후의 특허 침해로부터 보호받을 수 있다.

라. 기술 상세 정보

발명품과 관련된 기술적 세부 사항이 포함되어 있으며, 이는 주로 제목, 설명, 청구항, 요약, 그리고 도면 등으로 구성된다. 제목은 기술의 내용을 간결하게 표현하고, 요약에서는 그 기술의 주요 내용이 간략하게 기술된다. 설명 부분에서는 발명품을 사용하기 위한 상세한 지침과 함께, 해당 기술이 해결하려는 문제, 해결책, 실시 예, 그리고 달성된 효과 등이 상세히 기재되어 있다. 청구항에서는 출원자가 권리로 주장하고자 하는 핵심 사항들이 명시된다.

• 청구항의 중요성 : 청구항은 특허의 법적인 보호 범위를 정의한다. 이는 특허가 제3자에 의해 침해되었는지를 판단하는 기준이 되기 때문에 매우 중요하다.
• 도면 : 도면은 발명품이나 기술을 시각적으로 설명하며, 복잡한 기술이나 발명품을 이해하는 데 도움을 준다.

마. 특허 분류

특허 분류는 특허 문서를 보다 쉽게 분류하고 검색할 수 있도록 돕는 역할을 한다. 주로 국제특허분류(IPC) 코드가 사용되며, 해당 코드는 특허 문서에 명시되어 있다. 이를 통해 해당 기술이 어느 분류에 속하는지 확인할 수 있다. 일부 국가에서는 추가적인 특허 분류 시스템을 사용하며, 이러한 분류 시스템도 문서에 병기되기도 한다. 예를 들어, 미국의 USPC, 유럽의 ECLA, 독일의 DELKA, 그리고 일본의 F.I, F-Term 등이 이에 해당한다.

- 세부 분류 : IPC 외에도 특허에 대한 보다 세부적인 분류 시스템이 존재하며, 이를 통해 특정 기술 분야에 대한 정밀한 검색을 가능하게 만들어 준다.
- 분류 체계의 업데이트 : 기술이 발전함에 따라 특허의 분류 체계도 지속적으로 변하는데, 이는 새로운 기술 분야를 보다 정확하게 반영하기 위한 목적으로 수행된다.

2. 특허 문서의 구조

가. 기술 세부 정보

특허공보는 해당 기술의 상세한 설명과 핵심적인 요소를 제공한다. 기술의 세부 사항은 주로 설명서(description)와 청구항(claim) 부분에서 찾을 수 있다. 이러한 정보는 일정 시간이 지나면 대중에게 공개되어, 다양한 기술 분야에 대한 포괄적인 데이터베이스를 형성한다. 이는 특허 정보가 기술적인 지식을 포함하고 있다는 것을 의미한다.

나. 권리 관련 내용

- 특허 권리의 범위 : 등록된 특허의 경우, 청구항을 통해 해당 특허의 권리 범위를 확인할 수 있다. 또한 출원일을 기준으로 특허의 유효 기간을 계산할 수 있다. 특허 문서에서는 심사 과정, 등록 상태 등의 법적인 상황도 확

인할 수 있어, 특허 정보가 법적 권리와 관련된 데이터를 포함한다고 할 수 있다.

다. 경영 관련 내용

- 기술 개발 동향 및 경쟁 분석 : 출원인, 출원일 및 기술 내용 등의 정보를 활용하여 시간의 흐름에 따른 기술 개발의 트렌드를 분석하거나 특정 기업의 기술 출원 현황을 파악할 수 있다. 이를 통해 시장의 기술 동향이나 경쟁 기업의 연구 전략을 이해할 수도 있다. 따라서 특허 정보는 기업 경영의 관점에서도 가치 있는 데이터를 제공한다.

3. 선행기술조사 및 인용

가. 선행기술조사의 의미와 목적

특허에 있어서 선행기술조사는 특허성 여부 판단을 위해 출원된 발명과 동일하거나 유사한 종래의 기술이 존재하는지 여부를 조사·분석하여 심사관에게 전달하는 과정이다. 심사관은 선행기술조사 결과를 바탕으로 해당 발명이 신규성 있고 진보적인지, 그리고 그 발명이 특허로 등록될 만한 요건을 충족하는지 평가한다.

1) 신규성 또는 독창성 확인 : 선행기술조사를 통해 출원된 발명이 기존에 알려진 기술과 중복되는 것은 아닌지 여부를 확인한다. 신규성 또는 독창성이 부재하다면 해당 발명은 특허로 등록될 수 없다.
2) 진보성 평가 : 출원된 발명이 기존 기술에 비해 충분히 진보적인지도 특허 등록을 위한 평가 항목 중 하나이다. 만일 심사관이 진보성이 부족하다고 판단한다면 특허 등록을 위한 요건이 결여된 것으로 보아 특허가 부여되지 않는다.
3) 법적 안정성 확보 : 출원된 발명이 성공적으로 특허에 등록되면, 선행기술조사를 통해 해당 특허의 법적 안정성이 확보된다. 이는 나중에 발

생할 수 있는 특허 분쟁에서 유리한 위치를 차지하는 데 도움이 된다.

4) 기술 정보의 축적 : 선행기술조사 과정에서 수집된 정보는 다른 발명자와 연구자들에게 유용한 현행 기술에 대한 정보를 구축하는 데 도움을 준다.

나. 인용의 의미

특허 심사 과정에서 선행기술 문헌이 인용된다는 것은 출원된 발명이 해당 문헌에 기재된 기술과 관련이 있거나 유사하다는 것을 의미한다. 인용된 선행기술 문헌은 출원된 발명의 독창성과 진보성을 평가하는 데 중요한 근거가 된다. 특허 등록에 있어 선행기술 문헌을 인용하는 것이 가지는 의미는 다음과 같다.

1) 독창성 및 진보성의 근거 : 인용된 선행기술 문헌은 출원된 발명이 기존 기술에 비해 얼마나 독창적이고 진보적인지를 판단하는 데 사용된다. 현재의 기술 수준을 보여주는 지표로서 선행기술이 인용되고 나서, 해당 발명이 인용된 선행기술과 비교하여 무엇이 다르고 얼마나 진보적인지가 보다 명확하게 드러난다.
2) 특허 범위의 한계 설정 : 선행기술 문헌은 출원된 발명의 특허 범위를 명확하게 규정하는 역할을 한다. 선행기술 문헌과 비교하여 무엇이 같고 무엇이 다른지가 선행기술 문헌의 인용 과정에서 파악 가능하기 때문이다. 이는 나중에 발생할 수 있는 특허 분쟁에서 핵심적인 역할을 할 수 있다.
3) 기술 발전의 이해 : 인용된 선행기술 문헌은 해당 분야에서 당시 시점의 기술 발전 상황이나 트렌드를 잘 보여준다. 따라서 선행기술 문헌을 파악하면 시장의 흐름과 기술 발전의 방향을 이해하는 데 도움이 된다.
4) 연구 및 개발의 지향점 제공 : 인용된 문헌들은 다른 발명자나 연구자들에게 해당 분야에서 어떤 주제가 중요하고, 어떠한 특허가 기술적으로 중요한지, 표준적인 기술은 무엇인지, 어떤 방향으로 연구가 진행되고 있는지를 파악하는 데 도움이 된다.

이러한 선행기술조사와 인용 과정을 통해 특허 시스템은 새로운 발명이 기존 기술에 기반을 두면서도 충분히 독창적이고 진보적인지를 보장하고, 이를 통해 기술 발전을 촉진하며 혁신을 장려하게 된다.

4. 연구에서 활용한 특허 정보 데이터

본 연구에서 활용한 특허 정보 데이터는 특허정보활용서비스(KIPRIS)에서 제공하는 특허등록공보(서지) 데이터이다. 특허실용공보는 특허공개공보와 특허등록공보로 나누어지는데, 특허 등록 여부에 관계없이 출원된 내용을 모두 공개하는 특허공개공보를 사용하여 특허 등록이 성공적으로 이루어진 경우(즉 기술의 신규성 및 진보성이 인정된 경우)와 특허 등록에 성공하지 못한 경우(기술의 신규성 및 진보성이 결여되었다고 판단된 경우)에 대해 특허정보활용서비스에서 관리하고 있는 전 기간의 데이터를 모두 수집하였다.

특허공개공보는 초록, 청구항, 선행기술조사, 우선권, 특허코드, 관련인 등의 여러 가지 정보로 구성되어 있는데, 본 연구에서 주로 활용한 정보는 청구항과 선행기술조사이다.

청구항은 특허에 의해 부여된 보호의 범위를 법적으로 정의하는 부분으로, 특허로 보호되는 것이 정확히 무엇인지, 그렇지 않은 것은 무엇인지를 명확하게 설명한다. 즉, 청구항은 특허의 법적인 보호 범위를 정의하는데, 실제 소송에서 특허가 제3자에 의해 침해되었는지를 판단하는 기준이 되기 때문에 매우 중요하다.

특정 특허가 출원되면 특허 심사관은 한국특허기술진흥원 등을 통해 선행기술조사를 수행하게 된다. 선행기술조사는 앞서 설명하였듯, 특허성 여부를 판단하기 위해 출원된 발명과 동일하거나 유사한 종래의 기술이 존재하는지 여부를 조사·분석하여 심사관에게 전달하는 과정이다. 이 조사는 해당 발명이 신규성이 있고 진보적인지, 그리고 그 발명이 특허를 받을 수 있는 조건을 충족하는지를 평가하는 근거가 된다.

선행기술조사 보고서를 전달받은 심사관은 선행기술조사 보고서에 포함된 선행기술 문헌에 기재된 기술과 출원 신청한 특허의 내용 간의 관련성이

나 유사성을 발견하면 이를 인용한다. 즉, 선행기술조사를 마치게 되면 해당 특허와 유사하거나 기술적으로 관련된 특허(출원 또는 등록된)에 대한 출원 또는 등록 정보가 수집되고 수집된 기술이 해당 특허와 연관성이 있는지 심사관이 판단하는데, 그 과정을 인용이라고 한다. 인용은 단순한 관련성과 유사성을 의미하는 것이며, 어떤 특허에 인용된 문헌이 있다고 해서 이것이 바로 신규성 또는 진보성이 없다는 것을 의미하지는 않는다. 이러한 과정에 더해 추가적인 심사를 거쳐 심사관은 특허의 등록 여부를 결정한다. KIPRIS 내에서는 선행기술조사를 통해 관련 있는 것으로 제안된 문서 번호가 모두 데이터 내의 PriorTechnologyDocument에 등록되고, 그중 인용된 정보는 식별자 Y를 부여하는 것으로 해당 정보를 관리한다.

특허공개공보는 Abstract, AdministrativeProcess, Bibliographic, Claim, CPC, DesignatedCountry, Family, IPC, Priority, PriorTechnologyDocument, RelatedPerson, Rnd, Specification으로 구성되어 있다. 이 중 본 연구에서 사용하게 될 Claim과 PriorTechnologyDocument에 대한 설명은 다음과 같다.

가. Claim

특허 Claim은 특허에 의해 부여된 보호의 범위를 법적으로 정의하는 부분으로, 특허로 보호되는 것이 정확히 무엇인지, 그렇지 않은 것은 무엇인지를 명확하게 설명한다.

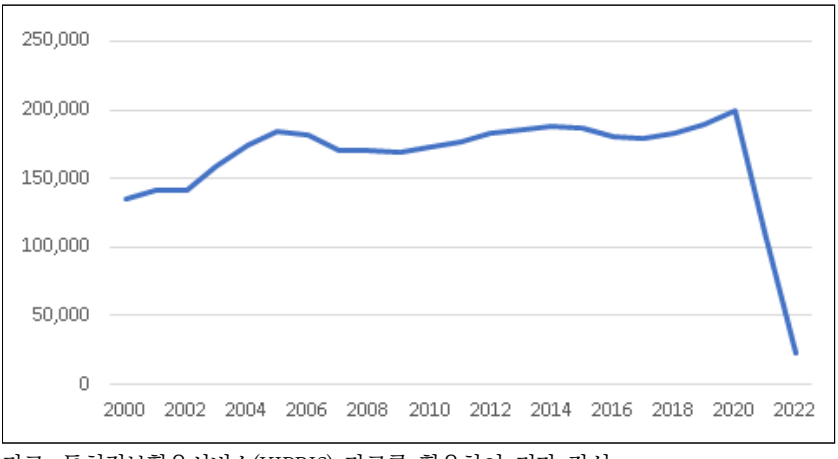
나. PriorTechnologyDocument

이 문서는 선행기술조사에 의하여 파악된 선행기술의 출원 또는 등록 번호를 포함하고 있다. 심사관에 의해 인용되었을 경우 Y의 식별자를 가지게 되며, 그렇지 않은 경우는 식별자 없이 출원 또는 등록 번호만을 표시한다.

[그림 5-1]은 본 장에서 분석에 사용할 수집된 특허 자료의 개수에 대한 연도별 통계를 그려놓은 것이다. 2023년 10월 기준 2022년에 출원된 특허가 아직 등록 심사를 마치지 않았거나 마친 특허라 해도 아직 특허의 모든 항목들이 시스템에 등재되어 있지 않기 때문에 수집 연도의 마지막 해인

[그림 5-1] 분석에서 사용할 특허의 연도별 개수

(단위 : 개수)



자료 : 특허정보활용서비스(KIPRIS) 자료를 활용하여 저자 작성.

2022년의 수치가 크게 감소하는 것을 확인할 수 있다. 2000년 이후 출원되어 시스템에 등록된 특허의 수는 매년 꾸준히 증가하는 추세를 보이고 있으며, 2012년 이후부터 2020년까지는 연평균 대략 18만 5,000 이상이다. 일반적으로 특허가 시스템에 등록되기 위해서는 특허가 공개되어야 하는데, 특허는 출원일로부터 18개월이 지나야 공개되어 시스템에 등재된다. 그 결과 2021년은 시스템상에서 확인되는 공개된 특허 건수가 11만 1,565로 전년도의 약 60%, 2022년은 2만 2,859로 2020년의 13% 정도에 불과하다. 따라서 데이터상에서의 편의 문제를 고려하여 자연어 처리 분석은 2000년부터 2020년까지 진행할 것이다.

제4절 분석 방법 및 결과

이번 장에서는 본 연구에 활용할 연구 방법론에 대해 설명한다. 앞서 특허정보 데이터에서 설명한 바와 같이, 특정 특허가 출원되면 특허 심사관은 한국특허기술진흥원 등을 통해 선행기술조사를 수행하게 된다. 선행기술조사

를 마치게 되면 해당 특허와 유사하거나 기술적으로 관련된 특허(출원 또는 등록된)에 대한 출원 또는 등록정보가 수집되며 수집된 기술이 해당 특허와 연관되어 있는지 여부를 심사관이 판단하게 되는데, 그 과정을 인용이라고 한다. 앞서 설명하였듯이, 선행기술조사를 통해 관련 있는 것으로 제안된 문서 번호는 모두 KIPRIS 데이터 내의 PriorTechnologyDocument에 등록되고, 그중 인용된 정보는 식별자 Y를 부여받는다.

1. 연구 방법 개요

우리가 분석할 정보는 권리의 범위를 확정하기 위해 claim 항목에 청구항으로 등재된 정보이다. 분석하고자 하는 정보의 청구항은 선행기술조사를 통해서 유관 기술로 보고되고 심사관을 통해 인용 여부가 확정된 기술의 청구항과 짝을 이루어 처리된다. 한 개의 출원 문서는 한 개 이상의 선행기술 문서와 짝을 이룬다. 따라서 분석하고자 하는 문서와 선행기술 문서는 1대 n의 관계를 이룬다.

앞서 언급한 바와 같이 본 연구에서는 신규성 및 진보성이 인정된 경우(Positive case, 특허 등록이 승인된 경우)와 신규성 및 진보성이 결여된 경우(Negative case, 특허 등록이 거절된 경우)를 모두 포함하는 특허등록공보를 사용하게 되므로 positive case와 negative case에 대한 정보를 얻을 수 있다. 이 짝을 가지고 분석하고자 하는 Claim이 기존 기술 대비 신규성 및 진보성이 인정되는 경우 1의 점수를, 신규성 및 진보성이 결여되어 있는 경우 0의 점수를 부여하여 이항 자료를 구축한 후 어떠한 특허가 1과 0 중 어떠한 값을 부여받을지를 예측하는 모형을 구축할 것이다. 그리고 구축이 완료된 모델을 이용하여 신규성 및 진보성을 (즉, 특허를 부여받게 될 가능성을) 0부터 1의 범위 내에서 예측하게 하였다.

가. BERT 모델

본 장에서 사용한 모형은 대표적인 언어모형 중 하나인 BERT(Bidirectional Encoder Representations from Transformers) 모형이다. BERT는 최근 각광

받고 있는 번역기 기반 모델 (Transformer based model) 중 하나로 처리할 수 있는 자연어의 총 길이가 제한되어 있는 반면 자연어 처리가 빠르고 사용이 쉽다는 장점을 가지고 있다. BERT를 사용하는 자연어 처리 과정은 다음과 같다.

1) 전처리

가) 토큰화(tockenization) : 문서에 있는 문장들이 토큰화, 즉 단어나 부분 단어로 나뉜다. BERT는 자체 토크나이저를 가지고 있어 입력을 토큰으로 나누고 모델이 uncase일 경우 소문자로 변환한다.

나) 특수 토큰 추가 : [CLS]는 문장의 시작에, [SEP]는 문장의 끝과 문장들 사이에 추가된다. [CLS]는 분류 작업에 사용되며, [SEP]는 다른 문장들을 구분하는 데 사용된다.

다) 패딩 및 절삭 : 토큰화된 문장들은 BERT 모델이 받아들일 수 있는 특정 길이로 패딩되거나 절삭된다.

2) 임베딩 생성

가) 입력 임베딩 : 토큰화된 문장들은 임베딩으로 변환되어, BERT 모델의 입력으로 사용될 수 있는 벡터 표현으로 바뀌게 된다.

나) 위치 임베딩 : BERT는 입력 임베딩에 위치한 임베딩도 추가하여 문장 내 각 단어의 위치를 파악할 수 있도록 해준다.

3) BERT 모델

임베딩은 여러 개의 트랜스포머 계층을 포함하는 BERT 모델로 공급된다. 각 계층은 입력을 변환하고 다음 계층으로 전달하며 그 과정 중에 단어 간의 복잡한 관계를 포착하게 된다. 최종 트랜스포머 계층의 출력은 하위 작업에 사용된다.

4) 분류 계층

두 문장의 연관성을 분류하기 위해 BERT 모델 위에 분류 계층이 추가된

다. [CLS] 토큰의 출력은 전체 입력 시퀀스에 대한 정보를 포함하고 있으며, 보통 분류 작업에 사용된다. 출력은 소프트맥스 계층을 통해 각 클래스에 대한 확률을 얻기 위해 전달된다.

5) 훈련 및 미세 조정(fine tuning)

관련성을 분류하는 작업인 경우, 훈련을 위해 문장 쌍과 이 쌍들에 해당하는 연관성 레이블이 포함된 레이블된 데이터 세트가 필요하다. BERT 모델은 모형을 미세 조정하면서 이 데이터 세트에서 예측된 연관성과 실제 연관성 간의 차이를 최소화하기 위해 가중치를 조정하면서 훈련을 진행한다.

6) 평가 및 추론

훈련 후, 모델은 별도의 검증 세트에서 평가되어 일반화가 잘 되는지 확인해야 한다. 추론을 위해 모델은 두 문장을 받아들여 BERT 모델과 분류 계층을 통해 처리한 후 연관성 분류를 출력한다.

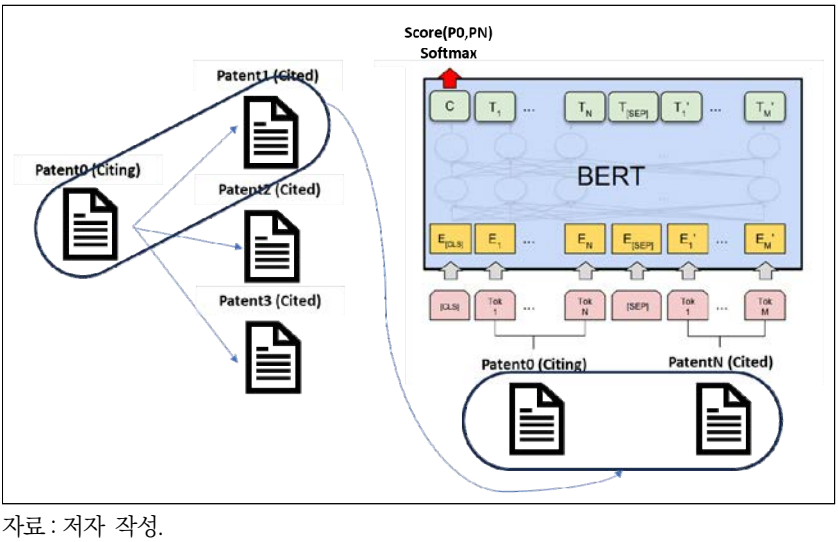
나. 신규성 분석을 위한 BERT 모델의 적용

일반적인 BERT를 사용하는 자연어 처리와 비교하여 본 연구에서 설계하는 모형의 특징은 다음과 같다. 우선 BERT 모델의 선택에 있어서 본 연구에서는 한국어 처리만을 위한 한글 BERT 모델과 더불어 다국어 모형(multi-linguistic model)을 사용하였다. 지금까지의 특허 신규성 분석 연구는 다음의 두 가지 문제를 안고 있다. 첫째로, 주로 영어로 쓰인 외국의 특허 정보를 분석한 연구가 이루어졌으며 반면 한글로 된 특허 정보의 분석이 거의 이루어지지 않았다. 둘째로, 언어 간의 비교 분석이 이루어지지 않은바, 외국어로 된 기술 선행 연구와 한글로 된 특허 정보 사이의 신규성을 체계적이고 일관된 분석 모형 안에서 분석할 수 없었다. 이 연구에서는 다국어 언어 모형을 선택하여 선행기술이 영어 또는 일본어와 같은 외국어로 이루어진 경우에도 동일한 모델 내에서 분석될 수 있도록 다국어 모형을 추가하였다.

두 문장의 연관성 분류를 위해서 앞서 훈련 및 미세조정 부분에서 언급하였듯이, 자연어 처리 중에 BERT를 사용했다. 이 구조는 두 문장의 연관성 훈련에 사용될 수 있을 뿐만 아니라, 미세조정을 통해 유사성 이외의 두 문장 간의 관계에 관한 다른 하위 작업에도 활용할 수 있다.

본 연구에서는 두 문장의 연관성 대신 신규성 분석 대상인 특허 정보(Citing, 그림 5-2에서 P0로 표시)를 첫 번째 문서로 입력하고 연관된 문서인 피인용 특허 정보(Cited, 그림 5-2에서 PN으로 표시)를 두 번째 문서로 입력하여, 첫 번째 문서가 두 번째 문서에 비해 신규성이 있으면 1인 케이스로, 아니라면 0인 케이스로 훈련하였다. 또한 1의 값을 얻은 경우에는, 문서 입력의 순서를 바꾸어 기존 선행기술인 첫 번째로 입력된 문서가 새롭게 신규성을 인정받은 두 번째로 입력된 문서에 비해 신규성이 떨어지는 경우로 변환하여 훈련하였다. 이러한 훈련 과정을 거치면 시스템은 평가하려고 하는 특허 정보가 기존 특허 정보(선행기술 문서로부터 추출된)에 비해 기술적 신규성 및 진보성이 존재할 경우 1에 가깝도록 값을 변환하고, 반대의 경우에는 0에 가깝도록 예측한다. 본 연구에서 사용되는 구조를 [그림5-2]에 도식화하였다.

[그림 5-2] 신규성 분석을 위한 기본 구조



자료 : 저자 작성.

2. 신규성 분석 결과

앞서 소개한 특허 자료와 방법론을 가지고 특허 신규성 분석에 대한 분석을 실시하였다. 한국어 BERT 모델은 SKTBrain에서 개발한 KoBERT 모델과 한국전자통신연구원(ETRI)에서 개발한 KorBERT 모델을 사용하여 성능을 비교하였다. 다국어 모델로는 Hugging Face에서 제공하는 BERT-base-multilingual-cased 모델을 사용하였다. BERT 모델은 최대 512토큰까지밖에 처리하지 못하기 때문에, 512 토큰을 상회하는 특허 정보는 불가피하게 절삭하였다. 훈련을 위한 Optimizer로는 Adam이 사용되었으며, 학습률은 2e-5로 설정되었다. 목적함수로는 Classification에 주로 활용되는 Cross Entropy Loss함수가 활용되었다. 데이터 세트는 랜덤하게 나누어져 80%가 훈련을 위해 사용되었고, 20%는 테스트를 위해 사용되었다.

비교 대상으로는 Risch et al.(2020)의 PatentMatch라는 시스템이 있으며, 해당 연구에서는 특허 신규성 분석에 있어 52~54%에 해당하는 정확도(accuracy)를 보여주었다. 이번 연구에서 달성한 한국어 특허 정보 간 정확도(accuracy) 와 F1 스코어 및 한국어-외국어 특허 정보 간 정확도는 〈표 5-2〉와 〈표 5-3〉에 정리되어 있다.

〈표 5-2〉의 결과와 같이 한국어 특허 정보 간의 신규성 분석에서는 Risch et al. (2020)의 PatentMatch 시스템이 달성한 예측력과 유사한 수준의 정확도성을 가진 예측력을 달성하였다. F1 값을 기준으로 모델 간의 예측력을 비교하면 한국어 특허 정보 간의 분석에서는 한글에 특화된 KoBERT 및 KorBERT 모형이 BERT-Multiligual 모델보다 정확도 및 F1에서 앞서는 성능을 보여주었다.

〈표 5-2〉 한국어 특허 정보 간의 신규성 분석

	정확도	F1 Score
KoBERT	53.01	55.08
KorBERT	57.49	60.24
BERT-Multilingual	50.17	54.91

자료 : 특허정보활용서비스(KIPRIS) 자료를 활용하여 저자 작성.

〈표 5-3〉 한국어-외국어 특허 정보 간의 신규성 분석

	정확도	F1 Score
KoBERT	34.56	38.71
KorBERT	35.88	39.92
BERT-Multilingual	37.12	40.04

자료 : 특허정보활용서비스(KIPRIS) 자료를 활용하여 저자 작성.

한국어-외국어 특허 정보 간의 신규성 분석은 한국어 정보 간의 신규성 분석보다 현저히 낮은 성능을 보였다. 이는 정확도와 F1 스코어 모두에서 공히 관찰되었다. 한편, 각 모형 간의 성능 비교에서는 Bert-Multiligual 모델이 KoBERT와 KorBERT 모델에 비해 근소하게 앞서는 성능을 보였는데, 다국어 모형으로서 Bert-Multiligual 모델이 가지는 장점을 고려한다면 실제 성능의 차이가 유의미하게 크지는 않은 것으로 보인다.

한국어-외국어 특허 정보 간의 분석에서 나타난 전반적으로 낮은 신규성 예측력은 대부분의 훈련 데이터가 한글 특허 데이터 간의 신규성 분석이었기 때문에 한글 특허 정보와 외국어로 된 특허 정보 간의 관계에 대한 학습이 충분히 이루어지지 않았기 때문인 것으로 생각된다. 다국어 모형의 성능이 다른 두 한국어 모형보다 크게 앞서지 않는 것은 대부분의 외국어 특허 정보가 미국 및 유럽 특허로 영어로 된 데이터인 만큼 KoBERT나 KorBERT의 훈련 시에도 충분한 영어 데이터가 학습되었기 때문인 것으로 추측된다.

제5절 피인용 횟수에 따른 특허 분석

1. 기초 통계

본 장에서는 특허에 대한 자연어 분석의 일환으로 피인용 횟수에 따라서 특허의 성격에 차이가 있는지를 자연어상의 특징을 가지고 살펴보고자 한다. 분석의 초점은 피인용 횟수에 따른 자연어상의 성격 차이지만, 기본적으

로 특허는 출원한 기업이나 개인이 속한, 혹은 특허에 해당하는 기술이 속한 산업의 특성에 따라서 피인용 횟수에 많은 차이가 발생할 수 있다. 예를 들어, 우리나라가 세계적인 경쟁력을 가지고 있고 기술을 선도하는 전지 · 전자 산업이나 자동차 산업, 디스플레이 산업 등에서는 등록 특허의 개수뿐만 아니라 개별 특허의 평균적인 피인용 횟수도 높을 가능성이 크다. 그러나 경쟁자가 적거나 혹은 우리나라에서 활성화되지 않은 산업의 경우, 특허의 출원과 등록 숫자만이 아니라 피인용 횟수도 적을 가능성이 있다. 따라서 이를 통제하기 위해서 등록된 특허의 개수가 많은 법인을 대상으로 하여 동일 법인이 출원한 특허 내에서 피인용 횟수별 특허의 특징을 비교하고자 한다.

이를 위해서 특허 정보 활용 서비스에 탑재된 등록 특허들에 대한 자료를 법인별로 취합하였다. 다만, 시기별로 특허의 유형, 종류, 특성, 양 등에서 차이가 존재할 수 있는바, 이를 통제하기 위해서 기술적 특성이 비교적 최근과 유사한 2001년부터 2020년까지의 특허만을 대상으로 하였다. 2021년 이후에 출원한 특허의 경우 아직 모든 특허가 등록 절차를 마치지 못하여 앞서 살펴보았듯이 다른 연도와 비교하여 개수 차이가 심하기 때문에 포함시키지 않았다.

특허 등록 수 상위 100개 법인 중에서 대학교 내 산학협력단 및 공공연구기관, 공공기관을 제외한 순수 민간 법인 5개 사를 선택하였으며, 이들 각 5개 사가 출원하여 등록된 특허를 인용 횟수에 따라 구분하였다. 이들 5개 사의 사명과 분석 대상 특허에 대한 기초 통계는 〈표 5-4〉에 정리되어 있다.

〈표 5-4〉 특허 등록 상위 5개 민간 법인의 피인용 횟수별 특허 수

	피인용 횟수 0	피인용 횟수 1-999	피인용 횟수 1,000이상	계
위니아	2,707	1,313	0	4,020
한국조선해양	2,064	1,186	0	3,250
엘에스일렉트릭	2,036	1,180	0	3,216
한화에어로스페이스	1,457	712	0	2,169
포스코	1,279	463	6	1,748

자료 : 특허정보활용서비스(KIPRIS) 자료를 활용하여 저자 작성.

〈표 5-4〉에서는 몇 가지 주목할 만한 특징을 발견할 수 있다. 첫째로 모든 5개 법인에서 등록된 특허 중 1회 이상 인용된 특허의 개수보다 한 번도 인용되지 않은 특허가 더 많다는 점이다. 피인용 횟수가 특허의 가치를 정확히 반영하는 지표가 아니라는 점은 본 보고서에서 수 차례 언급하였으나, 그럼에도 피인용 횟수가 있다는 점은 해당 특허가 신규성과 진보성에서 피인용 횟수가 없는 특허와 비교하여 더 나을 가능성을 시사한다. 왜냐하면 피인용되었다는 것은 해당 특허의 신규성이나 진보성에 기초하여 이후에 새롭게 개발된 특허가 있다는 것을 의미하며 이후의 기술 개발에서 참고나 시작점이 되었음을 의미하기 때문이다. 우리나라에 등록된 특허의 절대 다수가 피인용 횟수가 0이라는 점을 고려하면, 등록 특허의 수가 많은 법인들이 상대적으로 가치 있는 특허를 보다 높은 비율로 생산했다 할 수 있겠으나 여전히 많은 특허가 피인용 수 기준으로 높은 가치를 가지지 못하고 있다.

이러한 현상이 나타난 원인은 우리나라의 작은 시장 규모 때문이다. 예를 들어, 우리나라에서 가장 규모가 크며 세계적으로도 특허 출원 및 등록 건수가 많은 삼성전자나 LG전자 등은 모두 등록 특허 수 기준으로 상위 5개 법인에 들지 못했다. 이는 이들 기업들이 특허를 많이 생산하지 않아서가 아니라, 대부분의 특허를 미국이나 유럽 시장에 출원하고 한국에는 출원하지 않기 때문이다. 미국이나 유럽의 시장이 보다 크기 때문에 해당 국가에서 등록된 특허들은 상용화의 측면에서 높은 가치를 가진 특허일 것이며 따라서 피인용 횟수도 높을 가능성이 크나, 우리나라에 등록된 특허가 아니기 때문에 특허 정보 활용 서비스에서 찾아볼 수 없다.

물론 그렇다고 국내 대기업이 국내에 특허를 출원하지 않는 것은 아니나 보다 상품 가치가 높거나 중요한 특허는 미국이나 유럽에 등록하는 것이 현실이다. 따라서 여기에서 분석 대상으로 선정된 5개 법인은 국내에서 주로 영업하거나 혹은 국내에 출원한 특허가 많은 경우에 해당하는 사업체들이라 할 수 있다.

다른 한 가지 주목할 만한 점은, 포스코의 경우 1,000회 이상 피인용된 특허가 존재한다는 점이다. 다수의 인용된 특허가 한두 자리 수, 많으면 세 자리 피인용 횟수를 가지는 반면, 포스코의 경우 특정 6개의 특허가 압도적으로 많은 피인용 횟수를 가짐을 알 수 있다. 따라서 이들 6개 특허에 대해서

는 따로 하위 집단으로 묶어 분석하여 다른 인용된 특허와의 차이도 아울러 살펴보고자 한다.

2. 분석 방법 및 결과

분석은 일곱 가지 지표를 계산하는 것으로 진행하였다. 첫째 지표는 한 문서 내에 포함된 문장의 수이다. 문장은 의미 전달을 위한 단위이다. 따라서 한 문서에 많은 문장이 포함되어 있다면 그 문서는 더 많은 의미를 담고 있을 가능성이 높다. 특허에 얼마나 많은 의미가 담겨 있는지는 해당 특허의 특성을 살펴보는 데 있어 중요할 것이다.

둘째 지표는 한 문장 안에 포함된 단어의 수를 계산한 것이다. 한 문서 내에 포함된 문장의 수가 의미 전달의 숫자라면, 개별 문장 안에 평균적으로 얼마나 많은 단어가 포함되어 있는지는 해당 문장의 복잡성이나 단순성을 대변하는 지표이다. 많은 단어로 구성된 문장일수록 더 복잡한 문장이며 더 복잡한 의미를 담고 있다고 할 수 있다.

셋째로는 평균적인 단어의 길이를 측정한 지표이다. 특허는 필연적으로 전문적인 용어가 많이 사용될 수밖에 없는데, 일반적으로 전문용어는 일상어보다 길이가 길게 마련이다. 즉, 평균적인 단어의 길이가 길면 길수록 전문적인 지식을 보다 많이 담고 있을 가능성이 높다 할 수 있다.

넷째 지표는 한 문서 내에 포함된 어휘의 종류이다. 신규성이 높은 특허는 기존에 없던 새로운 지식이나 기술을 설명하고 있기 때문에 필연적으로 어휘의 종류가 다양해질 수밖에 없다. 이미 알려진 기술이나 지식은 해당 대상을 지목하는 어휘나 표현이 관용적으로 정착해 있기 때문에 서술이 길지 않을 뿐만 아니라 사용하는 어휘의 양도 감소하게 마련이다. 그러나 신규성이 높은 문서는 다양한 어휘로 새로운 내용을 설명하고 있을 가능성이 높다.

다섯째 지표는 3음절을 초과하는 어휘의 수이다. 앞서 단어의 길이에서 언급하였듯이 전문적인 지식을 설명하거나 표현하는 어휘는 일상어와 비교하여 긴 음절을 사용할 가능성이 높다. 물론 간단한 기술로 특허의 신규성과 진보성을 쉽게 풀어서 담아낼 수 있다면 좋겠으나 기존의 특허에서 진일보한 기술이나 지식을 담기 위해서는 필연적으로 어휘의 종류도 다양해져야

할 뿐만 아니라 어휘의 구성도 길어질 수밖에 없는 것이 일반적인 패턴이다.

여섯째 지표는 한 문서 내에 포함된 불용어(不用語)의 수를 세어 평균을 내는 방법이다. 불용어란 문장에서 의미에 큰 영향을 주지 않는 단어 혹은 색인 단어로 의미가 없는 단어를 지칭한다. 관사, 전치사, 조사, 접속사 등을 그 예로 들 수 있다. 불용어가 많이 사용된 문서일수록 문서 전체에서 중요도가 낮거나 의미에 영향을 미치지 않는 어휘의 비중이 높다는 것을 의미한다.

마지막으로는 SMOG(Simple Measure Of Gobbledygook) 지수를 사용하였다. SMOG 지수란 문서의 가독성을 측정하는 지표로 해당 문서를 이해하기 위해 필요한 교육 수준을 추정할 때 사용한다. 분석 텍스트를 시작, 중간, 끝 세 부분으로 나눈 후, 각각의 부분에서 10개의 문장을 추출하여 3음절 이상의 단어를 대상으로 다음절어의 개수를 세어 문서의 수준이나 가독성 등을 유추하는 지표이다.

특허 등록 수를 기준으로 뽑은 5개 사의 특허에 대해서 전술한 일곱 가지 지표를 각각 계산할 것이다. 앞서 언급하였듯이 특허의 특성은 업종이나 회사의 규모, 해외 진출 여부 등 다양한 요인에 의해서 영향을 받을 수 있으므로, 본 연구에서는 이를 통제하기 위해서 동일한 회사의 특허끼리 비교하였다. 특히 피인용 횟수가 0인 특허와 피인용된 특허 간의 자연어상의 차이를 체계적으로 분석하고자 특허를 피인용된 적이 한 번도 없는 특허, 피인용된 적이 있는 특허 두 집단으로 분류하였다. 만일 두 집단 간의 자연어 분석에서 일관된 방향의 차이가 관찰된다면 자연어를 기준으로 특허의 가치를 산정하거나 추정하는 것이 가능하다. 그러나 일관된 차이가 관찰되지 않는다면 특허의 가치는 앞선 일곱 가지 지표로는 관측 불가능하다 결론 내릴 수 있을 것이다.

본 분석에서 살펴보고자 하는 또 다른 것은, 실제 특허의 특성이 업종이나 기업에 따라서 차이를 보이는가 하는 점이다. 특허 등록 수가 많은 다섯 개 사를 꼽은 이유는 이들 기업이 출원하여 등록한 특허의 수가 많기 때문에 일부 특허에서 나타날 수 있는 이상치(outlier)의 특성이 전체에서는 희석되어 기업이 출원하는 특허의 전반적인 특성들이 충분히 대변될 것이기 때문이다. 충분히 많은 수의 특허를 가지고 업종이나 기업 규모, 해외 진출 여

부 등이 다른 법인들의 특허상의 특성을 살펴봄으로써 특허의 내용이나 자연어상의 성격에서 실제 업종이나 기업의 특성이 반영되어 차별점이 있는지를 살펴보는 것 역시 가능하다.

분석 결과가 정리된 〈표 5-5〉를 통해 몇 가지 사실을 확인할 수 있다. 첫째로 일곱 개의 지표 대부분에서 인용된 특허와 그렇지 않은 특허 간의 뚜렷한 차이를 관찰하기 어렵다. 전반적인 경향에 있어서는 차이가 크지 않으나 피인용된 특허들이 인용되지 않은 특허들과 비교했을 때 불용어의 수를 측정한 기준을 제외한 나머지 지표에서 대체로 크다. 하지만 이를 유의미한 차이라고 하기에는 어려워 보인다.

둘째로 피인용된 특허 가운데서도 1,000회 이상 피인용된 6개의 특허는 피인용 횟수가 적거나 한 번도 피인용되지 않은 특허들과 비교하여 한 문서 내에 포함된 평균적인 문장의 수는 많으나, 3음절을 초과하는 어휘의 수는 적으며, 불용어의 수는 많고, SMOG가 확연히 낮다. 그러나 이러한 차이가 철강 산업이라는 산업적 특성 차이에서 기인하는 것인지, 혹은 실제 피인용 횟수가 아주 많은 특허 전반에서 나타나는 것인지는 향후 다른 산업의 표본을 이용한 추가적인 분석이 필요하다 하겠다.

셋째로 피인용 횟수 여부에 따른 자연어상의 성격 차이보다는 업체 간의 자연어에서의 특성 차이가 더욱 크게 관찰된다. 예를 들어, 한국조선해양과 한화에어로스페이스를 보면 해당 법인이 출원 특허들을 피인용 횟수로 구분한 특허 집단 간의 차이는 크지 않지만, 이들 두 법인 간에는 거의 모든 지표에서 큰 차이를 발견할 수 있다.

따라서 자연어상으로 관찰되는 특허의 차이에 영향을 미치는 주요 요인은 업종과 같은 산업적 특성이 큰 반면 특허의 가치나 피인용 횟수는 큰 영향을 미치지 않는다고 할 수 있다. 다만, 피인용 횟수가 일정 수준을 초과하는 경우에는 자연어상에서 뚜렷한 차이가 나타날 가능성을 배제할 수는 없음도 아울러 확인했다.

〈표 5-5〉 피인용 횟수에 따른 특허의 자연어 분석

		피인용 횟수 1회 이상							피인용 횟수 0회						
		Sents[1]	SentLengths[2]	Words[3]	Word Vars[4]	Comp Words[5]	Stopwords[6]	SMOG[7]	Sents	SentLengths	Words	Word Vars	Comp Words	Stopwords	SMOG
위니아		6.9	10.1	3.4	38.4	32.4	11.9	15.5	7.1	10.3	3.3	39.6	35.2	13.8	15.8
한국조선해양		7.5	12.3	3.3	43.4	37.4	16.5	15.9	7.4	12.4	3.3	42.1	36.4	14.4	15.8
엘에스일렉트릭		6.6	11.6	3.2	37.0	30.5	14.5	15.4	6.3	12.6	3.1	35.4	30.1	14.9	15.6
한화에어로스페이스		8.3	10.5	3.1	37.7	34.3	15.1	14.7	7.6	11.0	3.0	35.5	32.0	13.0	14.9
포스코	전 체	7.3	9.4	3.1	32.1	23.0	8.7	13.3	6.7	10.1	3.2	31.4	24.2	9.1	14.0
	1,000회 이상	9.2	9.3	3.0	30.0	20.2	10.2	11.6							

주: 1) Sents : 한 문서 내에 포함된 문장의 수
2) SentsLentghs : 한 문장 안에 포함된 단어의 수
3) Words : 평균적인 단어의 길이
4) Word Vars : 한 문서 내에 포함된 어휘의 종류
5) Comp Words : 3음절을 초과하는 어휘의 수
6) Stopwords : 한 문서 내에 포함된 불용어의 수
7) SMOG Index : 6-Sixth grade~17-Graduate grade

자료 : 특허정보활용서비스(KIPRIS) 자료를 활용하여 저자 작성.

제6절 고용 변화와 특허의 자연어적 특성

앞선 절에서 자연어상의 특성을 살펴보기 위해서 일곱 가지의 지표를 사용하였다. 여기서는 동일한 지표를 가지고 법인의 고용 변화와 특허의 자연어적 특성 간에 어떠한 관계가 있는지를 살펴보고자 한다. 이를 위해서 분석 대상 기간인 2001년부터 2020년 사이에 비율 측면에서 고용이 많이 감소한 곳과 많이 증가한 곳을 꼽아 이들 기업이 첫 5년 간 등록을 승인받은 특허와 마지막 5년 간 출원하여 등록한 특허 간의 차이를 살펴보고자 한다.

그러나 등록된 특허의 개수가 충분치 않으면 분석에서 유의미한 결과를 얻을 수 없다. 따라서 2001년부터 2020년 사이에 등록된 특허의 개수가 가장 많은 상위 100개 법인을 추린 후, 이 중에서 앞선 절과 마찬가지로 공공기관과 연구기관, 대학과 대학의 산학협력단을 제외시키고 남은 민간기업들을 대상으로 고용 변화를 조사하였다. 여기서 2001년에 존속하지 않은 법인이나 2015년 이후에 사라진 법인은 고용원 수를 0으로 보지 않고 표본에서 제외하였다. 이는 0에서부터 증가율을 계산하는 것이 불가능하거나 0으로 변한 고용률을 살펴보는 것이 무의미할 뿐만 아니라, 이들은 무조건적인 고용 감소나 증가만 관찰되기 때문이다. 또한 첫 5년 간 존속하지 않은 기업은 당연히 등록한 특허도 없을 것이므로 자연어 분석 대상이 되는 특허도 찾기 어렵기 때문이다.

한편 앞서 특허의 자연어적 특성에서 업종 차이가 크게 나타났기 때문에 고용 증가율과 감소율 기준으로 각각 상위 40개 사를 꼽은 후 이들 중 업종이 유사한 곳들만 추려서 분석하였다. 등록된 특허들의 분석 기간은 2001년부터 2005년까지의 첫 5년과 2016년부터 2020년까지이다.

그 결과 해당 기간 고용 증가율이 높았던 기계 · 장치 산업 법인으로 만도, 현대트랜시스, 한화시스템이 꼽혔으며, 고용이 비교적 크게 감소한 것으로 나타난 법인으로는 두산중공업, 현대위아, 삼성중공업, 케이씨가 꼽혔다. 이들 기업 중 일부는 인수합병이나 기업의 분할 혹은 매각 등이 있었기에 고용원 수의 변화율이 크게 나타났으나, 이러한 변화 역시 특허의 속성에

영향을 미칠 수 있기에 이러한 이벤트를 통제하지 않고 법인을 선발하였다. 이들 기업들의 구체적인 고용 변화율에 대한 수치는 생략하고, 특허와 관련된 기초 통계를 살펴보면 다음과 같다.

우선 2005년까지 기업들이 출원한 특허의 수가 최근 5년에 비해 크게 적다는 것을 알 수 있다. 따라서 2001년부터 2005년까지 만도, 두산중공업, 현대위아의 경우 등록된 특허가 전무(全無)한 것으로 나타나 고용 증가율 상위 법인의 경우 현대트랜시스와 한화시스템의 특허만이, 고용 감소율 상위 법인의 경우 삼성중공업과 케이씨의 특허만 분석 가능하였다.

등록된 특허의 증가율을 보면 고용 증가율 상위 법인에서의 특허 등록 건수 증가가 16배에 가까운 것으로 계산되나, 고용 감소율이 큰 법인에서는 특허 등록 건수가 30배 넘게 증가한 것으로 조사되었다. 따라서 기업의 규모가 커지고 고용원 수가 많아져야만 특허 출원 및 등록 건수가 늘어난다고 결론 내릴 수 없음을 확인할 수 있다.

첫 번째 분석은 각 법인 단위로 기간별 특허의 특성에 차이가 있는지를 살펴보았다. 이 분석에서는 시간 불변적인 각종 법인 단위 효과가 모두 통제되어 있으므로 고용 변화와 등록된 특허 간의 특성 차이를 비교적 명확하게 살펴볼 수 있다.

분석 결과를 정리한 〈표 5-7〉을 살펴보면, 고용 변화율에 따른 집단 분류

〈표 5-6〉 고용 변화율이 큰 7개 사에 대한 특허 기초 통계

		2001~2005 등록 특허 수	2016~2020 등록 특허 수	전체 특허 수
고용 증가율 상위 법인	만도	0	221	221
	현대트랜시스	2	499	501
	한화시스템	87	701	788
고용 감소율 상위 법인	두산중공업	0	469	469
	현대위아	0	190	190
	삼성중공업	4	740	744
	케이씨	49	198	247
합 계		142	3,018	3,160

자료 : 특허정보활용서비스(KIPRIS) 자료를 활용하여 저자 작성.

〈표 5-7〉 고용 변화율과 특허의 자연어 속성에 대한 분석 1

		2001~2005							2015~2020						
		Sents	SentLengths	Words	Word Vars	Comp Words	Stopwords	SMOG	Sents	SentLengths	Words	Word Vars	Comp Words	Stopwords	SMOG
고용 증가	만도	-	-	-	-	-	-	-	7.3	11.8	3.2	41.5	35.2	12.8	15.7
	현대트랜시스	5.5	13.3	3.5	44.0	35.0	17.5	17.5	6.1	12.1	3.4	38.7	30.2	13.5	15.8
	한화시스템	9.2	12.5	3.2	48.0	45.1	23.7	15.8	7.1	14.2	3.1	40.4	32.9	17.8	15.4
고용 감소	두산중공업	-	-	-	-	-	-	-	8.4	10.9	3.1	43.0	37.1	15.3	15.1
	현대위아	-	-	-	-	-	-	-	6.3	11.9	3.4	37.5	30.8	13.0	15.8
	삼성중공업	13.3	8.9	3.3	63.3	51.3	27.5	14.3	6.6	10.2	3.2	34.1	27.9	8.9	14.9
	케이씨	9.6	12.5	3.2	52.4	49.5	20.5	16.1	8.3	11.8	3.1	44.7	39.9	15.4	15.7
기간별 전체 평균		9.4	12.4	3.2	49.9	46.7	22.6	15.9	7.1	11.9	3.2	39.2	32.4	13.7	15.3

자료 : 특허정보활용서비스(KIPRIS) 자료를 활용하여 저자 작성.

〈표 5-8〉 고용 변화율과 특허의 자연어 속성에 대한 분석 2

2001~2005 (만도, 현대트랜시스, 한화시스템) + 2015~2020 (두산중공업, 현대위아, 삼성중공업, 케이씨)							2015~2020 (만도, 현대트랜시스, 한화시스템) + 2001~2005 (두산중공업, 현대위아, 삼성중공업, 케이씨)						
Sents	SentLengths	Words	Word Vars	Comp Words	Stopwords	SMOG	Sents	SentLengths	Words	Word Vars	Comp Words	Stopwords	SMOG
7.4	10.9	3.2	38.9	33.1	12.7	15.2	6.9	13.0	3.2	40.5	32.9	15.7	15.6
전체 특허 수 : 1,636개							전체 특허 수 : 1,474개						

자료 : 특허정보활용서비스(KIPRIS) 자료를 활용하여 저자 작성.

에서 나타나는 특성보다는 시기별로 나타나는 특성 차이가 더욱 크다는 것을 알 수 있다. 공통적으로 한 문서 내에 포함된 어휘의 개수(WordVars)나 3음절을 초과하는 어휘의 개수(CompWords)에 있어서 큰 감소가 있었음이 두 집단에서 공히 확인되며, 불용어의 수 역시 감소하였다.

하지만 고용 변화율에 따른 집단 간에도 차이가 존재하는 지수도 있다. 예를 들어, 불용어의 수(Stopwords) 지표는 항상 고용이 많은 쪽, 즉 초반기에는 고용 감소율이 큰 집단에서, 그리고 후반기에는 고용 증가율이 큰 집단에서 크게 나타났다. 다시 말해서 고용 규모가 커짐에 따라 특허에서 평균적으로 등장하는 불용어의 개수가 늘어남이 확인되었다.

한편 집단을 크게 두 유형으로 분류하여 고용이 많았던 시기의 법인들을 하나로 묶고 고용이 적었던 시기의 법인들을 묶어서 살펴본 비교 연구에서는 앞서 언급한 불용어의 개수 차이와 한 문장 안에 포함된 단어 수의 차이 두 가지 지표에서 뚜렷하게 고용이 많았던 시기에 높은 값이 나타났음을 확인할 수 있다.

〈표 5-7〉과 〈표 5-8〉은 고용 변화와 자연어에서 나타나는 특성 간에 뚜렷한 관계를 찾기 어려움을 확인시켜 주고 있다. 특허의 특성을 결정짓는 주요 요소는 인력이나 기업의 규모보다는 업종이나 시기임을 확인할 수 있었다. 그러나 본 연구는 소수의 기업만을 대상으로 진행하였던 데다, 상대적으로 가치 있는 특허들은 주로 미국이나 유럽에 등록되므로, 같은 분석을 유럽이나 미국 자료를 가지고 수행하였을 때도 동일한 결과가 나타난다고 할 수는 없으며, 따라서 제한적인 일반화만이 가능하다고 결론지을 수 있다.

제7절 소 결

자연어 처리는 대량의 텍스트 데이터를 효과적이고 효율적으로 처리하는 방법을 제공하였으며, 특허 정보 분석에 있어서도 그 활용이 점점 늘어나는 추세에 있다. 그중 특허 신규성 예측은 특허의 대상이 되는 기술이 새로운 특허로 인정받을 만한 새로운 기술인지 아닌지를 기계를 통해 구분해 내는

기술이라고 정의할 수 있다. 지금까지의 신규성 분석 연구는 영어로 된 특허 데이터를 중심으로 이루어져 한글로 된 국내 특허에 적용하기 힘들었을 뿐만 아니라 서로 다른 언어로 작성된 특허 정보 간의 비교는 이루어지지 못했다.

본 연구에서는 한글로 된 특허 정보를 대상으로 신규성을 분석하였을 뿐 아니라, 다국어 언어모형을 활용하여 한글로 된 특허 정보와 외국어로 된 특허 간의 신규성 분석을 수행하여, 신규성 분석의 범위를 넓히고 그 활용도를 높이고자 하였다.

분석 결과에 의하면, 정확성을 기준으로 동일한 언어인 한국어 특허 정보 간의 신규성 분석에서는 57.49%의 예측력을, 한국어 – 외국어 특허 정보 간의 신규성 분석에서는 37.12%의 성능을 보여주었다. 또한 다언어 모형이 한국어-외국어 특허 정보 간의 신규성 분석에서 큰 성능 개선을 보여주지 않은 것으로 나타났다. 이러한 결과는 한국어 모형의 성능이 좋다는 것으로 해석하기보다는 이미 한국어 모형에서 영어에 대한 학습을 충분히 하고 있기에 다언어 모형이 추가적인 성능 개선을 내지 못한 것으로 파악하는 것이 바람직할 것이다. 반면 한국어-한국어 간의 신규성 분석에서는 한국어 모형들의 성능이 다언어 모형보다 앞서있음이 확인되어, 한국어로 된 자연어 모형을 개발할 근거를 제시한다고 볼 수 있다. 한편 앞으로 신규성 분석에 있어서 최신 자연어 처리 모델인 초거대언어모형을 적극적으로 활용하는 등 이 분야에 더 많은 연구가 필요한 것으로 판단된다.

특허의 가치를 대리하는 지표로서 피인용 횟수에 따라서 특허의 초록에서 자연어의 특성을 포착할 수 있는지에 대해 수행한 연구에서는 업종 간에는 뚜렷한 차이를 발견하였으나 피인용 횟수에 따른 차이는 비교적 크지 않게 나타났다. 따라서 특허의 가치를 평가함에 있어 경제학자들이 겪는 어려움을 자연어 분석을 통해 해결하기도 쉽지 않아 보인다. 다만, 이러한 분석 결과는 특허의 초록에 대한 분석에서만 얻어진 것이며, 특허의 본문이나 다른 자연어 자료와 결합시키는 경우에는 다른 결과를 얻을 수도 있다.

지난 20년간 고용이 크게 증가했거나 감소한 기계 · 장치 · 설비 업체들에 대해서 특허의 특성을 비교해 보았다. 이를 통해 자연어 분석으로 관찰 가능한 특허의 특징들이 고용 변화와 어떠한 관계가 있을지를 살펴보고자 하였

다. 분석 결과에 따르면, 특허의 속성은 시기별로 큰 차이를 보이는 것이 비교적 뚜렷하게 관찰되었다. 그러나 고용 변화와 관련하여는 일부 지표에서는 뚜렷한 차이가 관찰되었으나, 그러한 차이의 일반화에 대해서는 확신할 수 없는 결과를 얻었다. 다만, 이 분석 역시 초록만이 아닌 본문이나 다른 자연어까지 결합시킨다면 다른 결과를 얻을 가능성도 배제할 수 없을 것이다.

제 6 장
결론 및 정책적 시사점

지속적인 경제 성장을 위해서 필요한 여러 요건 중 하나는 연구개발을 통해 꾸준히 경제 전반의 생산성을 제고하는 것이다. 일반적으로 경제가 성장함에 따라 경제 전체의 성장률은 떨어진다. 또한 소득 수준이 높아지면서 출산율이 떨어져 여러 요인으로 경제성장률은 둔화되게 마련이다. 따라서 고소득 국가의 정부는 연구개발 활동을 장려함으로써 저출산에 따른 성장 패널티를 극복하고 높은 경제성장률을 유지하여 꾸준히 양질의 일자리를 창출하려고 노력한다. 우리나라 역시 국가연구개발 지원사업이나 연구소 전담부서 사업과 같이 연구개발 활동을 장려하기 위한 다양한 정책을 펼치고 있으며, 신속 심사 제도를 통해 특허 출원을 장려하고 중소기업의 특허 심사료를 지원함으로써 소규모 사업체의 연구개발 활동을 간접적으로 독려하고 있다.

본 보고서는 이러한 연구개발 활동이 기업 단위에서 실제 매출을 비롯한 경영 지표나 종사자 1인당으로 측정한 생산성, 고용지표 등에 어떠한 영향을 미치는지, 그리고 정부가 펼치는 연구개발 활동의 성과는 어떠한지를 평가해 보았다. 아울러 특허를 분석하는 도구로서 자연어 분석을 사용하여 특허의 가치 평가나 특허 산출에 따른 고용효과 등을 시도해 보았다.

연구의 결과를 요약하면 다음과 같다.

연구개발 활동이 실제 효과를 내기 위해서는 일정한 수준 이상의 규모여야 한다는 점을 확인했다. 임계점 이하의 연구개발 활동은 어떠한 지표로도

생산성이나 매출, 고용과 유의미한 관계를 찾을 수 없었다. 기업의 연구개발 활동과 생산성 간의 관계는 측정 지표가 무엇이냐에 따라 상이한 결과를 보여주었다. 이는 연구개발 활동이 일시적으로 증가하는 것보다는 장기적인 관점에서 이루어지는 것이 중요할 가능성을 시사한다. 더군다나 연구개발 활동은 반드시 일정 수준 이상의 산출물을 보장하는 투자가 아니라는 점에서 연구개발 활동에 대한 투자가 이루어지는 시점에서는 투자가 비용에 가까운 성격을 지닐 수 있기 때문에 생산성이나 매출에 즉각적인 양의 효과를 내지 않은 것으로 보인다. 한편 고용과 연구개발 활동 간의 강한 양의 상관관계를 확인할 수 있었다. 이것은 기업 규모가 커짐에 따라 연구개발 활동을 수행할 여력이 생기고, 시장 점유율 유지와 매출 확대를 위해 연구개발 활동을 전개할 유인이 발생한다고 해석할 수도 있고, 다른 한편으로는 연구개발 활동을 위해 고용하는 인력은 기존의 생산 인력이나 사무직 인력과는 학력이나 숙련 요건 등이 다르기 때문에 신규채용을 유발한다고 할 수도 있다.

국가연구개발 지원사업의 경제적 효과를 평가한 결과는 다음과 같다. 첫째로, 국가연구개발 지원사업에 참여한 기업은 참여 이후 등록 특허의 수가 증가하지만 그 효과는 장기간 지속되지는 않는 것으로 보인다. 특히나 사업 참여 이전부터 등록 특허의 수가 증가 추세에 있다는 사실까지 고려한다면, 국가연구개발 지원사업이 연구개발 활동의 산출물을 증가시키는 효과는 크지 않을 가능성도 시사한다. 둘째로, 당기순이익과 매출 등의 경영성과 지표는 사업 참여 이후에 오히려 악화되는 것으로 나타났으나, 이러한 효과가 장기에도 계속 나타나는지는 추가적으로 긴 시계열을 확보한 분석이 필요할 것으로 보인다. 연구개발 활동은 투자 집행 시점에서는 비용과 같은 성격을 지니므로 일시적으로 경영 지표를 악화시킨 후, 이러한 악화된 지표의 영향이 이후 몇 년 동안 나타났을 가능성도 있기 때문이다. 셋째로 국가연구개발 지원사업 참여와 고용 변수 간에는 양의 상관관계가 관찰되었다. 따라서 국가연구개발 지원사업이 장기적으로 참여 기업의 규모를 키우고 생산성을 늘려 종사자 수도 늘렸을 가능성을 시사한다.

특허에 대한 자연어 분석은 세 가지 측면에서 이루어졌다. 첫째로, 자연어를 이용하여 특허의 신규성을 판별하는 작업을 수행하였다. 그 결과는 한국어 관련 자연어 모델이 한국어 특허 문서에 대해서는 다른 언어 모델보다

좋은 성능을 보였고, 또한 기존의 영어권 특허 문서를 대상으로 한 연구와 유사하거나 높은 수준의 성과를 보였다. 피인용 횟수에 따른 자연어상의 특허 특성의 차이를 살펴본 분석에서는 업종 간에 뚜렷한 차이는 나타났으나 피인용 횟수의 차이가 특허의 자연어 특성 차이로 드러나지는 않는 것으로 나타났다. 마지막으로 지난 20년간 고용률이 크게 증가한 법인과 크게 감소한 법인 간의 특허의 차이를 자연어로 분석한 결과에서는, 시기별 차이가 뚜렷하게 나타났으나, 고용이 적었던 시기와 고용이 크게 증가하거나 많았던 시기 간에 자연어상의 차이는 발견할 수 없었다.

이러한 연구 결과를 바탕으로 몇 가지 정책적 시사점을 찾을 수 있을 것이다.

첫째로, 연구개발 활동이 제대로 된 효과를 발휘하기 위해서는 일정 규모 이상의 투자가 필요함이 확인되었다. 따라서 연구개발 활동을 장려하기 위한 정책은 사업의 수를 늘리기보다는 개별 사업이나 정책의 예산 규모를 키우는 것에 주력할 필요가 있다. 국가연구개발 지원사업의 경우 개별 기업 입장에서는 개별 사업의 예산이 큰 편인바, 그로 인해서 특허 출원 성과 및 고용 성과 등에서 양호한 지표가 나타난 것으로 보인다.

이러한 연구개발의 속성은 기업 단위 혹은 국가 단위에서 연구개발에 투입하는 자원의 규모에 따른 '연구개발에 근거한 빈곤 함정(R&D-based poverty trap)'의 가능성을 시사한다. 즉, 저소득 국가나 생산성이 낮은 기업은 연구개발에 투자하는 자금의 규모가 작기 때문에 뚜렷한 생산성 증대 효과를 얻지 못하고, 그로 인해서 소득 수준이나 경영 성과가 계속해서 낮은 상태에 머물러 있을 수 있다. 반대로 고소득 국가나 규모가 큰 기업은 많은 자원을 연구개발 활동에 투입하여 뚜렷한 생산성의 증대와 앞선 기술에 따른 성장 효과를 누림으로써 기업 간, 나아가 국가 간 소득 격차가 더욱 확대될 수 있다. 이는 우리나라에서 향후 연구개발 활동을 펼칠 때 대규모 기업에의 집중 투자에 대한 근거가 될 수도 있으며, 한편으로는 중소기업에 대한 소수 집중 투자로의 정책 방향 전환을 모색할 여지도 동시에 제공한다.

둘째, 대기업과 중견기업의 경우, 일반적으로 연구개발 활동을 수행할 필요성을 충분히 가지고 있다. 따라서 정책 지원을 우량한 중소기업이 충분한 예산 지원을 받아 필요한 연구를 수행함에 따라 스케일업하여 대기업이나

중견기업으로 성장하도록 유도하는 방향으로 전개할 필요가 있다. 그 결과 성장한 중소기업은 자체적인 연구개발 수행 역량을 갖추게 되기 때문에 더 이상의 국가 지원을 필요로 하지 않고, 따라서 다른 우량한 중소기업에 다시 성장의 기회가 돌아갈 수 있다.

그러나 이 과정에서 피터팬 증후군처럼 중소기업이 충분히 성장할 수 있음에도 성장하지 않는 경우를 방지하기 위해서 중복 참여나 수혜에 대한 제한을 걸어두는 것도 필요할 수 있다.

셋째, 연구개발 활동과 고용 간에는 순환 구조가 존재한다. 제대로 된 연구개발 활동을 수행하기 위해서는 기업의 규모가 일정한 수준 이상이어야 하며, 또한 성공적인 연구개발 활동의 수행은 다시 고용을 늘릴 가능성이 분석 결과에서 확인되었다. 이는 기존의 사무직이나 생산직이 연구개발 인력으로 전환되는 것이 쉽지 않기 때문에 연구개발 활동을 전개하기 위해서는 전문적인 전담 연구개발 인력을 필요로 하기 때문이다. 따라서 양질의 일자리를 확보하기 위해서는 기업의 연구개발 활동을 장려하고 촉진할 필요가 있다.

한편 특허의 출원 수가 정확한 생산성을 대변치 못한다는 연구 결과에도 주목할 필요가 있다. 과도한 특허 출원은 특허 심사료 등의 사회적 비용을 증가시키고 아울러 실효성 있는 연구개발 활동이 기존에 출원된 많은 양의 특허와 비교하여 높은 신규성이나 진보성을 가지는 것으로 판별될 가능성도 낮출 수 있다. 현재 우리나라의 특허 심사료는 낮은 편인바, 이로 인한 과도한 출원과 심사관의 업무 과중을 야기할 수 있다. 따라서 중소기업이나 일정 규모 이하의 기업에 대한 특허 출원료 지원은 계속하되, 전반적인 특허 출원이나 심사, 등록 비용을 올리는 것도 검토 가능하다.

마지막으로 특허에 대한 자연어, 특히 한국어 기반 자연어 연구의 필요성이다. 어떠한 특허에서 담고 있는 기술이 새로운 특허로 인정받을 만한 새로운 기술인지 아닌지를 판별하였을 때, 각각의 특허가 서로 다른 가치를 가지고 있는 것처럼 특허의 신규성 수준에서도 각 특허는 다른 정도의 신규성을 가지고 있을 것이다. 향후 특허 심사 과정에서 한국어로 된 자연어 모형을 구축한다면, 우선은 자연어로 해당 특허가 가진 신규성의 정도를 어느 정도 가늠할 수 있거나 정확하게 판별할 수 있게 되며, 따라서 특허의 가치를 산

정하는 한 기준으로서 향후 자연어로 측정한 신규성을 이용할 수 있을 것이다. 이는 이후 R&D 활동에 대한 연구에 있어서 특허의 가치 추정이나 연구개발 활동의 성과 평가, 사업 평가 등에 활용될 수 있을 것이다. 또한 정확한 신규성이나 생산성의 측정을 통해 특허나 연구개발 활동이 고용을 비롯한 경제 전반이나 사회에 미치는 효과에 대한 분석과 측정에서도 진일보한 자료나 측정 지표를 제공할 수 있을 것이다.

참고문헌

김정언(2006),『지적재산권강화가 기술혁신 및 생산성에 미치는 효과분석 : IT 산업을 중심으로』, [KIIP] 한국지식재산연구원도서DB.

김정언 · 강성진(2007),「특허권강화와 특허출원변화의 기술혁신 및 생산성 파급효과 : 산업내 및 IT 산업의산업간파급효과를 중심으로」,『기술혁신연구』15(1), pp.145-173.

김진영(2012),「한국기술연구인력의 특허생산성」,『한국경제의 분석』18(2), pp.1-44.

서환주 · 이영수(2005),「특허권강화는 국가간 성장격차를 확대시키는가? : 특허권강화와 기술혁신간의 상관관계를 중심으로」,『[KIEP] East Asian Economic Review』9(2), pp.119-143.

양성준 · 김동현(2021),「AI 취약직업 : 직업적특성과 공간적특성을 중심으로」,『한국측량학회학술대회자료집』2021(11), pp.38-43.

오동현 · 김소영(2015),「우리나라 중소기업의 생산성, 기술혁신, 기술추격 및 특허권의 상관관계에 관한 연구」,『지식재산연구』10(1), pp.225-255.

이경곤(2018),「The Effects of R & D Intensity in the Industrial Sectors on Wage Distribution in the United States」,『대한경영학회지』31(10), pp.1827-1848.

장선미(2020),「연구개발과 특허가 생산성에 미치는 영향에 관한 연구」,『통상정보연구』22(4), pp.375-393.

장지연 · 정준호 · 심지환(2022),『기술변화가숙련과 노동수요에 미치는 영향 : 온라인 구인공고와 특허출원데이터분석을 중심으로』, 한국노동연구원.

정성철 · 윤문섭 · 장진규(2004),「특허와 기술혁신 및 경제발전의 상관관계」,『정책연구』, pp.1-114.

조상섭 · 정동진 · 정해식(2003),「OECD 25 개국특허자료를 이용한 지식파

급(Knowledge Spillover) 효과분석」, 한국무역학회세미나 및 토론회, pp.57-71.

Alain, G., and Y. Bengio(2016), "Understanding intermediate layers using linear classifier probes," arXiv preprint arXiv:1610.01644.

Alesina, A., M. Battisti, and J. Zeira(2018), "Technology and labor regulations : theory and evidence," *Journal of Economic Growth* 23, pp.41-78.

Altuntas, Serkan, Turkay Dereli and Andrew Kusiak(2015), "Analysis of patent documents with weighted association rules", *Technological Forecasting and Social Change* 92, pp.249-262.

An, X., J. Li, S. Xu, L. Chen, and W. Sun(2021), "An improved patent similarity measurement based on entities and semantic relations," *Journal of Informetrics* 15(2), Article 101135.

Arras, L., F. Horn, G. Montavon, K.-R. Müller, and W. Samek(2017), ""What is relevant in a text document?" : An interpretable machine learning approach," *PLoS One* 12(8), e0181142.

Arts, S., B. Cassiman, and J. C. Gomez(2018), "Text matching to measure patent similarity," *Strategic Management Journal* 39(1), pp.62-84.

Arts, S., J. Hou, and J. C. Gomez(2021), "Natural language processing to identify the creation and impact of new technologies in patent text : Code, data, and new measures," *Research Policy* 50(2), Article 104144.

Autor, D., A. Salomons, and B. Seegmiller(2023), "Patenting with the stars : Where are technology leaders leading the labor market?."

Bahdanau, D., K. Cho, and Y. Bengio(2014), "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv : 1409.0473.

Bamakan, Seyed Mojtaba Hosseini, Najmeh Faregh, and Ahad ZareRavasan(2021), "Di-ANFIS : an integrated blockchain-IoT-big

data-enabled framework for evaluating service supply chain performance", *Journal of Computational Design and Engineering* 8(2), pp.676-690,

Blei, David M., Andrew Y. Ng, Michael I. Jordan(2003), "Latent Dirichlet Allocation", *The Journal of Machine Learning Research* 3, pp.993-1022.

Castelvecchi, D.(2016), "Can we open the black box of AI?," *Nature News* 538(7623), 20.

Chen, Yi-Min, Yu-Ting Ni, Hsin-Hsien Liu, and Ying-Maw Teng(2015), "Information and rivalry-based perspectives on reactive patent litigation strategy", *Journal of Business Research* 64(4), pp.788-792.

Chiang, T.-J.(2011), "Defining patent scope by the novelty of the idea," *Wash. UL Rev* 89, p.1211.

Chikkamath, R., M. Endres, L. Bayyapu, and C. Hewel(2020), "An empirical study on patent novelty detection : A novel approach using machine learning and natural language processing," Paper presented at the 2020 Seventh International Conference on Social Networks Analysis, Management and Security(SNAMS).

Crampes, C., and C. Langinier(2002), "Litigation and settlement in patent infringement cases," *RAND Journal of Economics*, pp.258-274.

Danilevsky, M., K. Qian, R. Aharonov, Y. Katsis, B. Kawas, and P. Sen(2020), "A survey of the state of explainable AI for natural language processing," arXiv preprint arXiv : 2010.00711.

Danzer, A., C. Feuerbaum, and F. Gaessler(2020), "Labor supply and automation innovation," Max Planck Institute for Innovation & Competition Research Paper(20-09).

Dechezleprêtre, A., D. Hémous, M. Olsen, and C. Zanella(2019), "Automating labor : evidence from firm-level patent data," Available at SSRN 3508783.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova(2018), "Bert : Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv : 1810.04805.

Fall, Caspar J., Attila Törcsvári, Karim Benzineb, and G. Karetka(2004), "Automated categorization in the international patent classification", *ACM SIGIR Forum* 37(1), pp.10-25.

Freunek, M., and A. Bodmer(2021a), "BERT based freedom to operate patent analysis," arXiv preprint arXiv:2105.00817.

_____(2021b), "BERT based patent novelty search by training claims to their own description," arXiv preprint arXiv:2103.01126.

Gambardella, A.(2005), "Patents and the division of innovative labor," *Industrial and Corporate Change* 14(6), pp.1223-1233.

Gaur, M., K. Faldu, and A. Sheth(2021), "Semantics of the black-box : Can knowledge graphs help make deep learning systems more interpretable and explainable?," *IEEE Internet Computing* 25(1), pp.51-59.

Gerken, J. M., and M. G. Moehrle(2012), "A new instrument for technology monitoring : Novelty in patents measured by semantic patent analysis," *Scientometrics* 91(3), pp.645-670.

Grawe, Mattyws F., Claudia A. Martins, Andreia G. Bonfante(2017), *Automated Patent Classification Using Word Embedding*, Proceedings of the International Conference on Machine Learning and Applications(ICMLA), pp.408-411.

Hain, Daniel, Roman Jurowetzki, Tobias Buchmann, and Patrick Wolf(2021), Text-based Technological Signatures and Similarities : How to create them and what to do with them, working paper, arXiv:2003.12303v2.

Hasan, M. A., W. S. Spangler, T. Griffin, and A. Alba(2009), "Coa : Finding novel patents through text analysis," Paper presented at the Proceedings of the 15th ACM SIGKDD international conference on

Knowledge discovery and data mining.

Hepburn, Jason(2018), "Universal Language Model Fine-tuning for Patent Classification", Proceedings of the Australasian Language Technology Association Workshop 2018, pp.93-96.

Jang, H., Y. Jeong, and B. Yoon(2021), "TechWord : Development of a technology lexical database for structuring textual technology information based on natural language processing," *Expert Systems with Applications* 164, Article 114042.

Kaiser, U., H. C. Kongsted, and T. Rønde(2008), *Labor mobility and patenting activity*.

Kang, Seokho, Pilsung Kang, Taehoon Ko, Sungzoon Cho, Su-jin Rhee, Kyung-Sang Yu(2015), "An efficient and effective ensemble of support vector machines for anti-diabetic drug failure prediction", *Expert Systems with Applications* 42(9), pp.4265-4273.

Khattak, Akmal Saeed and Gerhard Heyer(2011), *Significance of Low Frequent Terms in Patent Classification using IPC Hierarchy*, IICS 2011, pp.239-250.

Kim, J., and S. Lee(2015), "Patent databases for innovation studies : A comparative analysis of USPTO, EPO, JPO and KIPO," *Technological Forecasting and Social Change* 92, pp.332-345.

Kim, Keungoui, Sungdo Jung, Junseok Hwang, and Ahreum Hong(2018), "A dynamic framework for analyzing technology standardisation using network analysis and game theory", *Technology Analysis & Strategic Management* 30, pp.540-555.

Kim, Mujin, Youngjin Park, and Janghyeok Yoon(2016), "Generating patent development maps for technology monitoring using semantic patent-topic analysis", *Computers & Industrial Engineering* 98, pp.289-299.

Kim, S., and B. Yoon(2021), "Patent infringement analysis using a text mining technique based on SAO structure," *Computers in Industry*

125, Article 103379.

Kim, Y.(2014), "Convolutional Neural Networks for Sentence Classification," Paper presented at the Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Kong, Aejing, Yuan Zhou, Yufei Liu, and Lan Xue(2017), "Using the data mining method to assess the innovation gap : A case of industrial robotics in a catching-up country", *Technological Forecasting and Social Change* 119, pp.80-97.

Krestel, R., R. Chikkamath, C. Hewel, and J. Risch(2021), "A survey on deep learning for patent analysis," *World Patent Information* 65, Article 102035.

Kyebambe, Moses Ntanda, Ge Cheng, Yunqing Huang, Chunhui He, and Zhenyu Zhang(2017), "Forecasting emerging technologies : A supervised learning approach through patent analysis", *Technological Forecasting and Social Change* 125, pp.236-244.

Kyung-gon, L.(2018), "The Effects of R & D Intensity in the Industrial Sectors on Wage Distribution in the United States,"『대한경영학회지』31(10), pp.1827-1848.

Larkey, Leah(1999), Patent Search and Classification System, Digital Libraries 99 - The Fourth ACM Conference on Digital Libraries (Berkeley, CA, Aug. 11-14 1999) ACM Press, pp.79-87.

Lee, C., B. Kang, and J. Shin(2015), "Novelty-focused patent mapping for technology opportunity analysis," *Technological Forecasting and Social Change* 90, pp.355-365.

Lee, C., B. Song, and Y. Park(2013), "How to assess patent infringement risks : A semantic patent claim analysis using dependency relationships," *Technology Analysis & Strategic Management* 25(1), pp.23-38.

Lee, Changyong, Ohjin Kwon, Myeongjung Kim, and Daeil Kwon(2018), "Early identification of emerging technologies : A machine learning approach using multiple patent indicators", *Technological*

Forecasting and Social Change 127, pp.291-303.

Lee, Jieh-Sheng and Jieh Hsiang(2020), "Patent classification by fine-tuning BERT language model", *World Patent Information* 61.

Leydesdorff, L., D. Kushnir, and I. Rafols(2014), "Interactive overlay maps for US patent(USPTO) data based on International Patent Classification (IPC)," *Scientometrics* 98 (3), pp.1583-1599.

Li, M., X. Ming, L. He, M. Zheng, and Z. Xu(2015), "A TRIZ-based trimming method for patent design around," *Computer-Aided Design* 62, pp.20-30.

Li, Rongrong, Xuefeng Wang, Yuqin Liu, and Shuo Zhang(2021), "Improved Technology Similarity Measurement in the Medical Field based on Subject-Action-Object Semantic Structure : A Case Study of Alzheimer's Disease", *IEEE Transactions on Engineering Management* 99, pp.1-14.

Li, Shaobo, Jie Hu, Yuxin Cui, Jiangjun Hu(2018), "DeepPatent : patent classification with convolutional neural networks and word embedding" *Scientometrics* 117(2), pp.721-744,

Li, Yahyong and John Shawe-Taylor(2007), "Advanced learning algorithms for cross-language patent retrieval and classification", *Information Processing and Management* 43(5), pp.1183-1199.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, … and V. Stoyanov (2019), "Roberta : A robustly optimized bert pretraining approach," arXiv preprint arXiv:1907.11692.

Loshchilov, I., and F. Hutter(2017), "Decoupled weight decay regularization," arXiv preprint arXiv:1711.05101.

Lu, Y., X. Xiong, W. Zhang, J. Liu, and R. Zhao(2020), "Research on classification and similarity of patent citation based on deep learning," *Scientometrics* 123(2), pp.813-839.

Marco, A. C., J. D. Sarnoff, and A. Charles(2019), "Patent claims and patent scope," *Research Policy* 48(9), Article 103790.

Mikolov, T., K. Chen, G. Corrado, and J. Dean(2013), "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781.

Moro, Sergio, Paulo Cortez · Paulo Rita(2017), "A framework for increasing the value of predictive data-driven models by enriching problem domain characterization with novel features", *Neural Computing and Applications* 28, pp.1515-1523.

Mou, L., R. Men, G. Li, Y. Xu, L. Zhang, R. Yan, and Z. Jin(2015), "Natural language inference by tree-based convolution and heuristic matching," arXiv preprint arXiv : 1512.08422.

Mullenbach, J., S. Wiegreffe, J. Duke, J. Sun, and J. Eisenstein(2018), "Explainable prediction of medical codes from clinical text," arXiv preprint arXiv:1802.05695.

Plantec, Q., P. Le Masson, and B. Weil(2021), "Impact of knowledge search practices on the originality of inventions : A study in the oil & gas industry through dynamic patent analysis," *Technological Forecasting and Social Change* 168, Article 120782.

Ranaei, Samira and Arho Suominen(2017), Using Machine Learning Approaches to Identify Emergence : Case of Vehicle Related Patent Data, 2017 Portland International Conference on Management of Engineering and Technology (PICMET). 1-8, CorpusID:43201462.

Risch, J., N. Alder, C. Hewel, and R. Krestel(2020), "Patentmatch : a dataset for matching patent claims & prior art," arXiv preprint arXiv:2012.13919.

Romer, P. M.(1990), "Are nonconvexities important for understanding growth?," In : National Bureau of Economic Research Cambridge, Mass., USA.

Seo, M., A. Kembhavi, A. Farhadi, and H. Hajishirzi(2016), "Bidirectional attention flow for machine comprehension," arXiv preprint arXiv:1611.01603.

Shibayama, S., D. Yin, and K. Matsumoto(2021), "Measuring novelty in science with word embedding," *PLoS One* 16(7), e0254034.

Small, H., K. W. Boyack, and R. Klavans(2014), "Identifying emerging topics in science and technology," *Research Policy* 43(8), pp.1450-1467.

Smolny, Werner(2000), "Sources of productivity growth : an empirical analysis with German sectoral data", *Applied Economics* 32, issue 3, pp.305-314.

Strumsky, D., and J. Lobo(2015), "Identifying the sources of technological novelty in the process of invention," *Research Policy* 44(8), pp.1445-1461.

Sun, Z., C. Fan, Q. Han, X. Sun, Y. Meng, F. Wu, and J. Li(2020), "Self-explaining structures improve nlp models," arXiv preprint arXiv:2012.01786.

Suominen, Arho, Hannes Toivanen, and Marko Seppanen(2017), "Firms' knowledge profiles : Mapping patent data with unsupervised learning", *Technological Forecasting and Social Change* 115, pp.131-142.

Tan, Shicheng, Shu Zhao, and Yanping Zhang(2022), Coherence-Based Distributed Document Representation Learning for Scientific Documents.

Tatikonda, M. V., and S. R. Rosenthal(2000), "Technology novelty, project complexity, and product development project execution success : A deeper look at task uncertainty in product innovation," IEEE *Transactions on Engineering Management* 47(1), pp.74-87.

Valentino, M., M. Thayaparan, and A. Freitas(2020), "Explainable natural language reasoning via conceptual unification," arXiv preprint arXiv:2009.14539.

Vapnik, Vladimir N.(2013), *The nature of statistical learning theory*. Springer Science & Business Media.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and I. Polosukhin(2017), "Attention is all you need," Advances in Neural Information Processing Systems 30.

Venugopalan, Subhashini and Varun Kumar Rai(2015), "Topic based classification and pattern identification in patents", *Technological Forecasting and Social Change* 94, pp.236-250.

Verhoeven, D., J. Bakker, and R. Veugelers(2016), "Measuring technological novelty with patent-based indicators," *Research Policy* 45(3), pp.707-723.

Wang, J., and Y.-J. Chen(2019), "A novelty detection patent mining approach for analyzing technological opportunities," *Advanced Engineering Informatics* 42, Article 100941.

Xu, Jian, Jiapeng Mu, and Gaorong Chen(2020), "A multi-view similarity measure framework for trouble ticket mining", *Data & Knowledge Engineering* 127, pp.1-17.

Yang, Yann-Jy, Jiann-Chyau Hwang(2020), "Recent Development Trend of Blockchain Technologies : A Patent Analysis", *International Journal of Electronic Commerce Studies* 11(1), pp.1-12.

Yoon, Byungun and Christopher L. Magee(2018), "Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction", *Technological Forecasting and Social Change* 132, pp.105-117.

Yuxin Chen, Jean Baptiste Bordes, and David Filliat(2017), An experimental comparison between NMF and LDA for active cross-situational object-word learning. In 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics, ICDL-EpiRob 2016.

Zanella, G., C. Z. Liu, and K.-K.-R. Choo(2021), "Understanding the trends in blockchain domain through an unsupervised systematic patent analysis," *IEEE Transactions on Engineering Management* 70, Issue 6.

◈ 執筆陣

• 방형준(한국노동연구원 연구위원)
• 곽도원(고려대학교 국제대학원 교수)
• 이가현(홍익대학교 경영대학 조교수)

자연어 기반 특허와 노동시장 성과 분석

▪ 발행연월일 | 2023년 12월 26일 인쇄
2023년 12월 29일 발행
▪ 발 행 인 | 허 재 준
▪ 발 행 처 | 한국노동연구원
3 0 1 4 7 세종특별자치시 시청대로 370
세종국책연구단지 경제정책동
☎ 대표 (044) 287-6080 Fax (044) 287-6089
▪ 조판 · 인쇄 | 고려씨엔피 (02) 2277-1508
▪ 등 록 일 자 | 1988년 9월 13일
▪ 등 록 번 호 | 제13-155호

 정가 6,000원

ISBN 979-11-260-0677-9

KLI
한국노동연구원

한국노동연구원

30147 세종특별자치시 시청대로 370 경제정책동
TEL : 044-287-6083 http://www.kli.re.kr

ISBN 979-11-260-0677-9