

## 데이터 기반 경제학 연구의 현황과 과제\*

최재성\*\* · 황윤재\*\*\*

### 논문초록

증거기반 정책에 관한 관심과 이를 위해 행정데이터를 활용하려는 노력은 세계적인 흐름이며, 정부가 보유한 행정데이터를 연계하고 이를 활용하여 인과관계를 규명하는 연구는 정책개발과 정책평가에 중요한 역할을 담당하고 있다. 한국에서도 증거기반 정책연구를 위한 행정데이터의 활용에 대한 논의는 과거부터 꾸준히 진행되었지만, 성과는 아직 미흡한 수준이다. 본 연구는 행정데이터를 활용한 국내 연구의 가능성을 파악하고 향후 정부와 학계의 협력을 위한 토대를 마련할 목적으로 진행되었다. 이를 위해 경제학 분야 4대 대표 학술지에 게재된 논문 정보를 수집한 후 이를 활용하여 행정데이터를 활용한 해외 연구의 동향을 검토하였다. 다음으로는 한국경제학회 회원이 제출한 12편의 연구계획서를 토대로 국내 행정데이터를 활용해서 수행할 수 있는 연구를 소개하고, 이에 필요한 행정데이터를 정리하였다. 본 연구는 정부가 증거기반 정책을 위한 행정데이터 공개 대상과 행정데이터 간 연계 작업의 우선순위를 결정하는 과정에서 유용하게 활용될 수 있을 것이다. 아울러 본 연구를 통해 제안된 행정데이터에 대한 수요가 반영되고 접근성이 개선된다면, 정책의제를 개발하고 정책평가를 수행하는 국내 실증연구가 빠르게 확대될 것으로 기대한다.

핵심 주제어: 행정데이터, 증거기반 정책

경제학문헌목록 주제분류: A1

투고 일자: 2024. 2. 22. 심사 및 수정 일자: 2024. 3. 13. 게재 확정 일자: 2024. 3. 20.

\* 본 논문은 저자들이 2023년 교육부의 의뢰로 진행한 '사회분야 데이터 기반 정책 연구의 현황과 과제' 최종보고서를 수정하고 발전시킨 것임을 밝힌다. 본 논문 작성 과정에서 자료 조사와 분석에 도움을 준 성균관대 박사 과정 서혜림 학생과 석사 과정 이도훈 학생에게 감사의 마음을 전한다. 아울러 긴박하게 요청을 드렸음에도 국내 행정데이터 활용 연구의 가능성과 지평을 넓히고자 연구계획서를 준비해주셨던 한국경제학회 회원들께 다시 한번 감사드린다. 이러한 노력이 한국 사회에서 데이터 기반 정책연구의 마중물이 되어 향후 정부와 학계의 협력이 확대되기를 기대한다.

\*\* 제1저자, 성균관대학교 글로벌경제학과 부교수, e-mail: jaesungc@skku.edu

\*\*\* 교신저자, 서울대학교 경제학부 교수, e-mail: whang@snu.ac.kr

## I. 서 론

증거기반 정책(evidence-based policy)에 관한 관심과 이를 위해 행정자료를 활용하려는 노력은 여러 선진국을 중심으로 빠르게 확대되고 있다. 미국 의회는 증거기반 정책을 추진하기 위해 2018년 “Foundations for Evidence-Based Policymaking Act”를 제정하였고, 미국 교육부는 수석 이코노미스트 자리를 신설하고 최고 데이터 책임자를 겸임하는 경제학자를 채용하여 관련 데이터를 정비하고 증거기반 정책의 토대를 구축하고 있다.

사회과학 내 여러 학문 분야에서 민간과 공공이 보유한 빅데이터를 활용한 실증 연구 또한 확대되고 있다. 특히 정부가 보유한 행정데이터를 연계하고 이를 활용하여 인과관계를 규명하는 연구는 정책개발과 정책평가에 중요한 역할을 담당하고 있다. 대표적으로 노르웨이, 스웨덴, 그리고 덴마크처럼 오랜 기간 복지 행정 시스템을 운영해온 유럽 국가들을 중심으로 조세 자료를 포함한 다양한 행정데이터를 연계하고 이를 활용한 연구가 활발히 수행되어왔다(유종성 외, 2020; 최재성, 2022).

미국에서는 연방 세금 자료를 사용해서 장기간에 걸친 세대 간 이동성(intergenerational mobility)을 분석하거나(Chetty et al., 2014), 소득과 자산 불평 등 추이를 검토하는 연구가 2010년 이전부터 진행되어왔다(Kopczuk et al., 2010; Piketty and Saez, 2014). 최근에는 소득세 자료에 대학 입시 자료인 SAT나 ACT 성적 그리고 개별 대학이 보유한 지원자 정보 등을 결합하여, 사회경제적 배경에 따른 교육 기회 및 노동시장 성과의 격차, 그리고 그러한 격차를 발생시키는 원인을 규명하는 연구도 수행되고 있다(Chetty et al., 2023).

우리나라 또한 관련 부처의 노력으로 건강보험DB나 고용보험DB 등이 정책연구에 일부 활용되고 있지만, 행정자료를 정책연구나 학술연구에 활용한 사례는 다른 선진국에 비해 부족한 수준이며, 자료의 적극적 연계 및 접근성 향상을 위한 추가적인 노력이 필요한 상황이다. 여러 기관이 보유한 다양한 행정데이터가 연계될 때 자료의 활용 가치가 한층 높아질 수 있음에도, 아직은 단일 기관이 보유한 데이터의 제공에 그치는 경우가 대부분이다.

사회보장위원회를 중심으로 행정데이터를 연계하는 노력이 진행 중이지만, 이 자료를 분석하기 위해서는 세종시 사회보장위원회에 직접 방문해야 하는 제약이 있다. 통계청이 제공하는 행정데이터 또한 서울과 지방에 설치된 소수의 거점 데이터

센터에 방문하고 상당한 비용을 지불해야 접근이 가능하므로, 개별 연구자가 장기간에 걸쳐 연구를 수행하기에는 부담이 크다. 이러한 이유로 여러 선행연구에서 지적한 바와 같이, 국내에서도 증거기반 정책연구를 위한 행정자료의 활용에 대한 논의는 과거부터 꾸준히 진행되었지만, 성과는 아직 미흡한 수준이다(김재진 외, 2020; 이영욱 외, 2020, 유종성 외, 2020)

본 연구에서는 해외 사례와 비교를 통해 국내 행정데이터의 접근성과 활용도를 높일 방안을 찾고, 이를 통해 증거기반 정책의 토대를 구축할 필요성을 제시하고자 한다. 이를 위해 행정데이터를 활용한 경제학 분야 해외 학술연구를 검토한다. 경제학 분야 최우수 국제학술지에 게재된 실증연구 중 행정데이터를 사용한 연구를 조사하고, 이들 연구에 활용된 행정데이터의 종류와 범위를 정리하여, 여러 나라에서 공통으로 활용 사례가 많은 데이터를 확인한다.

아울러 행정데이터에 대한 수요를 파악하기 위하여 한국경제학회 회원에게 의뢰하여 제출받은 12편의 연구계획을 소개한다. 2023년 하반기에 교육부는 사회 난제 해결을 위해 데이터 기반 정책연구를 확대하고자 한국경제학회와 협업을 제안했으며, 이를 위해 한국경제학회는 경제학 내 여러 세부 분야에 걸쳐 행정데이터를 활용한 연구주제를 발굴하는 작업을 진행하였다. 특히 행정데이터 간 연계를 통해 자료의 가치와 정책적 활용도를 높일 방안을 확인하는 작업에 중점을 두었다. 이를 바탕으로 증거기반 정책연구 활성화를 위한 정책 방안을 제안하고, 행정데이터 중 경제학계의 관심과 활용 가능성이 큰 자료를 확인하고, 행정데이터 간 연계의 우선순위를 제안한다. 더불어 연구자들이 행정데이터에 접근할 기회를 넓히고, 이를 이용한 연구를 활성화할 방안을 제안한다.

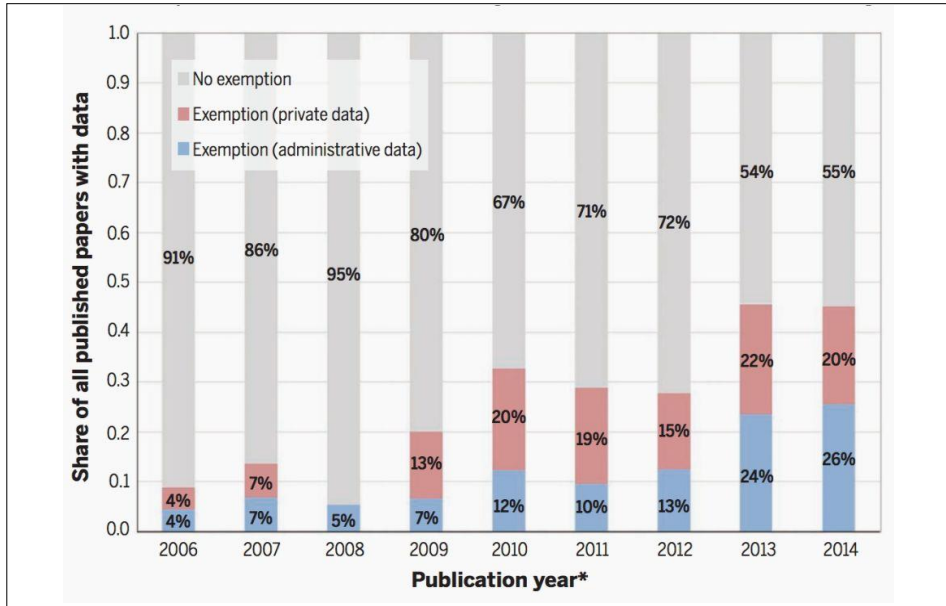
## II. 경제학 연구에서 행정데이터의 역할과 중요성

경제학 분야에서 행정 데이터를 사용한 연구는 2010년 이후 매우 빠르게 증가하고 있다. 한 연구에 따르면 경제학의 대표 저널인 American Economic Review에 게재된 실증연구 중 행정데이터를 사용한 연구의 비중은 〈Figure 1〉에 나타난 것처럼, 2006년 4%에서 2014년 26%로 증가하였다(Einav and Levin, 2014).

연구에 활용된 다양한 행정 데이터가 존재하지만, 그중에서도 국세청이 보유한 소득 및 재산에 관한 정보, 교육부가 제공하는 학업성취도 평가 및 교원 자료, 법

무부의 범죄 기록, 그리고 복지 제도의 운용을 위해 축적해온 다양한 개인 수준의 정보가 연구에 적극적으로 활용되고 있다.

〈Figure 1〉 Trends in Research Using Administrative Data in the American Economic Review, 2006-2014



Source: Einav and Levin, 2014.

미국에는 행정데이터를 활용하여 연구를 수행하는 학자들이 매우 많은데, 그중에서도 하버드 대학의 Raj Chetty가 지난 10여 년간 “The Equality of Opportunity Project”를 통해 교육과 사회이동성 분야에서 뛰어난 업적을 쌓아왔다. 최근에는 대학별 입시 데이터와 국세청 소득 정보를 연계하여 대학 입학과 관련된 실증연구를 수행하였고(Chetty et al., 2023), 이 연구 결과는 미국뿐만 아니라 여러 나라의 언론을 통해 소개된 바 있다.

행정데이터를 사용한 Chetty 연구팀의 주요 연구 결과 중에서는 취학 전 단계인 유치원의 질이 장기적으로 인생의 중요한 지표들에 영향을 미침을 보이거나, 소득 계층에 따라 고등교육의 접근성이 얼마나 차이를 보이는지를 규명한 연구가 있다. 그뿐만 아니라 입시 관련 성적이 유사함에도 가구 소득 수준에 따라 명문대 진학에 큰 차이가 있음을 보이고, 성장한 지역에 따른 기회와 성과 격차를 확인하고, 그러

한 격차를 발생시키는 이유를 행정데이터를 활용해서 규명한 연구도 있다.

이외에도 교육 분야에서 행정데이터가 활용된 사례는 여러 국가에서 다수의 연구를 쉽게 찾아볼 수 있다. 예를 들어 미시간주의 행정데이터와 미시간대학의 입학 사정 자료를 결합한 연구에서는 지원자의 성적이 비슷함에도, 왜 소득이 낮은 가정의 뛰어난 아이는 소득이 높은 가정의 아이보다 명문대에 진학하는 비율이 낮은지를 살펴보았다(Dynarski et al., 2016). 이 연구를 통해 연구자들은 입시 초기 단계에 학비를 전액 보장해주는 경우, 미시간대에 지원하는 비율과 최종 등록하는 비율이 큰 폭으로 상승하는 것을 발견하였다. 저자들은 이 연구의 결과를 바탕으로 불확실성과 위험회피 성향이 저소득층 학생의 대학 진학 결정에 중요한 역할을 한다는 점을 강조하며 정책적 시사점을 제공하고 있다.

대학에서 어떤 전공을 선택하고, 어떤 대학에서 어떤 동료들과 공부하는지가 대학 졸업 후 노동시장 성과에 영향을 주는지는 여러 나라에서 보편적으로 관심이 많은 주제다. 이러한 질문에 답하고자 노르웨이의 행정데이터를 사용한 한 연구에서는 고등교육 관련 정보(출신대학, 전공)와 졸업 후 소득 정보를 활용하였다. 이러한 행정데이터를 사용하여 분석한 결과를 보면, 전공에 따른 소득 차이는 크지만, 전공 차이를 고려하면 출신 대학의 순위에 따른 차이는 미미한 수준으로 나타났다(Kirkeboen et al., 2021).

이처럼 교육 분야뿐만 아니라 다양한 영역에서, 사회 문제를 진단하고 해결하기 위하여 행정데이터를 활용하는 연구들은 여러 나라에서 활발히 진행되어왔다. 최근 들어 그러한 연구 경향은 더 확대되고 있으며, 경제학 실증연구에서 행정데이터의 역할은 점점 더 그 중요성이 높아지고 있다.

### Ⅲ. 행정데이터를 활용한 해외 연구 검토

경제학을 포함한 다양한 사회과학 분야에서 행정데이터를 포함하는 빅데이터를 활용한 실증연구가 빠르게 증가하고 있다. 경제학 분야에 초점을 맞춰 최근의 추세를 좀 더 자세히 살펴보기 위해 경제학 분야를 대표하는 최우수 4대 학술지(American Economic Review, Quarterly Journal of Economics, Journal of Political Economy, Econometrica)에 게재된 논문을 대상으로 경제학 연구에 행정데이터를 활용한 사례를 직접 검토해보았다. 이 학술지들은 경제학 전체 분야에 걸쳐 순수

이론 논문부터 이론과 실증분석을 병행하거나, 실증분석 위주의 논문까지 다양한 스펙트럼의 논문을 폭넓게 수록하고 있다.

해외 경제학 연구에서 행정데이터를 활용한 연구를 식별하기 위해 본 연구에서는 2011년에서 2022년까지 이들 4대 학술지에 게재된 3,278편의 논문에 대한 메타 정보를 EBSCOhost 연구 플랫폼에서 웹스크래핑을 통해 수집하였다. 이 플랫폼에서는 논문의 기본적인 서지정보(제목, 저자, 게재 시점, 권호 등)와 함께 논문의 초록, 그리고 경제학 내 세부 분야를 표시하는 JEL 분류 기호, 그리고 해당 연구와 관련된 지역 정보를 확보할 수 있다. 다만 2024년 2월 현재 EBSCOhost에서 2023년 하반기에 게재된 논문 정보를 아직 제공하지 않는 학술지들이 있어, 본 연구에서는 분석 대상이 되는 학술지의 출판 시점을 2022년 말까지로 설정하였다.

분석 과정에서는 학술지에 실렸지만, 정규 논문이 아닌 경우는 제외하였다. AEA Papers and Proceedings에 실린 논문은 모두 제외하였으며, 각 학술지에 실린 이전 논문에 대한 Comments나 Reply 또한 제외하였다. 정제 과정을 거친 후 본 연구의 분석 대상이 된 논문의 학술지별 분포를 살펴보면 American Economic Review(AER)에 게재된 논문이 1,369편으로 가장 많았고, Quarterly Journal of Economics(QJE)에 실린 논문이 501편, Journal of Political Economy(JPE)에 실린 논문이 616편, Econometrica(ECTA)에 실린 논문 792편이 포함되었다.

먼저 행정데이터를 사용한 논문의 특성을 파악하기 위해 2022년에 출간된 312편(AER 113편, QJE 48편, JPE 69편, ECTA 82편)의 논문은 5명의 연구자가 초록뿐만 아니라 본문까지 살펴보면서 행정데이터를 활용하였는지를 직접 분류하였다. 다만 2011년에서 2022년의 전체 논문을 모두 직접 검토하기에는 논문의 수가 많아 이번 연구에서는 논문의 제목과 초록에서 행정데이터의 활용을 보여주는 단어(administrative data, administrative record, tax record 등)가 등장하는지를 먼저 확인하였다. 만약 제목이나 초록에 행정데이터 관련 키워드가 존재해서 행정데이터를 사용했을 것으로 보이는 논문은 초록뿐만 아니라 논문의 데이터 관련 본문까지 직접 읽고 행정데이터 사용 여부를 점검하였다. 이 과정을 통해 “administrative unit”과 같이 행정 구역을 의미하는 표현 등이 “administrative”와 함께 등장한 경우는 제외할 수 있었다.

물론 이렇게 제목과 초록만을 활용해서 분류하는 방식으로는 실제로 행정데이터를 사용한 논문의 수를 과소 추정하게 된다는 한계가 있다. 하지만 제목과 초록에

기초한 분류 방식을 일관되게 사용했을 때 시간에 따라 행정데이터를 활용하는 실증연구가 어떠한 추이를 보이며 변해왔는지를 확인할 수 있다는 점에 의의가 있다.

아울러 이러한 한계를 부분적으로나마 보완하기 위하여 분석의 시작 시점인 2011년에 4대 학술지에 출간된 240편의 연구 또한 초록을 직접 읽고 본문의 데이터 관련 서술까지 더불어 살펴보았다. 이를 통해 2011년과 2022년에 경제학 분야 4대 학술지에 출간된 연구 중 행정데이터를 사용한 논문 수의 변화를 정확하게 가늠해보았다.

앞서 설명한 방식을 적용해서 분석한 결과를 보면 전체 3,278편의 논문 중에서 109편의 논문의 제목이나 초록에서 행정데이터 사용을 시사하는 키워드를 찾을 수 있었다. 아래에는 2011년부터 2022년 사이에 경제학 분야 4대 대표 학술지에 게재된 학술논문 중에서 제목과 초록에 행정데이터 관련 키워드가 포함된 논문을 대상으로 분석한 결과를 정리하였다.

## 1. 행정데이터 활용 논문 추이 분석

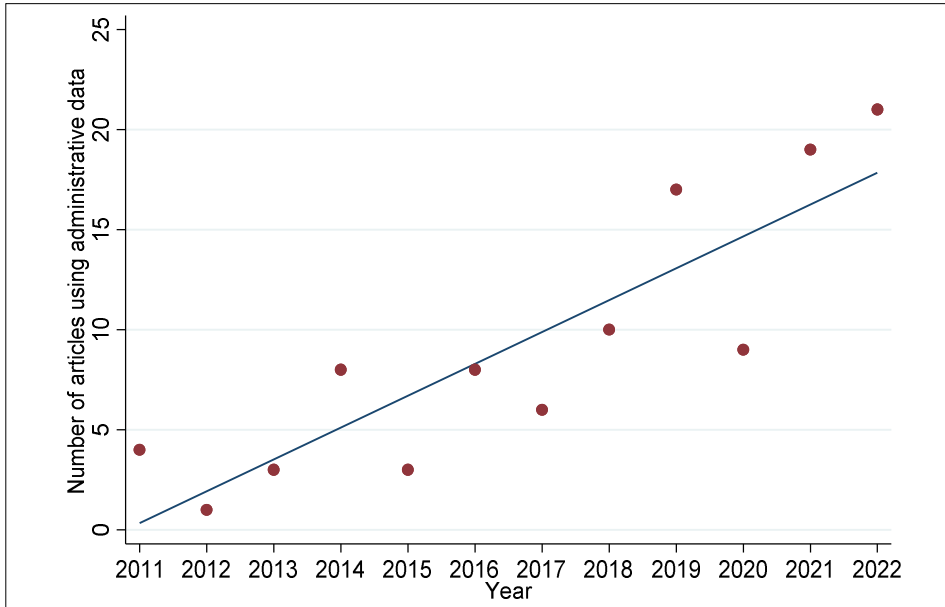
행정데이터를 사용한 연구의 연도별 분포는 <Figure 2>에 정리하였는데, 행정데이터를 활용한 학술논문의 수가 꾸준히 증가하는 추세를 확인할 수 있다. 다만, 이 결과는 제목과 초록에 행정데이터와 관련된 키워드가 등장하는지만을 사용해서 판별했다는 점에서 행정데이터를 활용한 연구를 과소 추정하고 있다. 2022년에 게재된 논문의 본문까지 직접 읽고 분류해보면, 행정데이터를 사용한 연구는 AER에서 18편, QJE에서 18편, JPE에서 8편, 그리고 ECTA에서 11편이 확인되어 총 55편에 달한다.

제목과 초록만을 활용해서 분류할 때 2022년에 게재된 논문 중 행정데이터 활용 논문의 수가 21편이라는 점을 고려하면, 실제로 본문까지 읽고 분류해서 확인된 55편이라는 수치는 행정데이터를 활용한 논문의 수는 다른 연도에도 <Figure 2>에 제시된 것보다 많을 것임을 시사한다. 2011년에 4대 대표 학술지에 게재된 논문을 대상으로 제목과 초록에 행정데이터 관련 키워드가 포함된 경우를 확인하는 방식으로 행정데이터를 사용한 4편의 연구를 찾을 수 있었으나, 본문의 데이터 관련 서술까지 포함해서 살펴보면 행정데이터를 사용한 10편의 연구를 찾을 수 있었다.

과거에는 행정데이터를 사용한 연구가 최근에 비해 상대적으로 많지 않았기 때문

에 이를 초록에서 언급하는 연구의 비중이 최근의 연구에 비해 높았을 것이라는 점을 고려하면, 행정데이터를 사용한 연구의 증가 추세는 〈Figure 2〉에서 보여지는 것보다 더 가파를 수 있다.

〈Figure 2〉 Trends in the Number of Papers by Year Using Administrative Data



## 2. 국가별 분석

다음으로 국가별 분포를 살펴보면, 미국과 유럽의 국가들에서 행정데이터를 활용한 연구가 활발하게 진행되고 있음을 볼 수 있다. 2011년에서 2022년까지 발간된 논문에 부여된 지리적 기술자를 이용해서 살펴보면, 제목과 초록을 통해 행정데이터를 사용한 것으로 분류된 109편 중 절반을 넘는 60편의 논문이 미국의 행정데이터를 사용한 것으로 나타났다. 다음으로는 덴마크의 행정데이터를 사용한 연구가 9편, 스웨덴의 행정데이터를 사용한 연구가 8편으로 그 뒤를 이었다. 독일의 행정데이터를 사용한 연구는 6편이었고, 또 다른 북유럽 국가인 노르웨이의 행정데이터를 사용한 연구는 5편이었다. 그다음으로 프랑스의 행정데이터를 사용한 연구가 4편 확인되었다. 2편의 연구가 확인된 국가는 그리스, 칠레, 캐나다와 파키스탄이었고,



그 밖에도 행정데이터를 사용한 한 편의 연구가 확인된 국가가 13개국 존재하였다. 이렇게도 해당 기간에 한국의 행정데이터를 사용하여 4대 학술지에 게재된 연구는 없었다.

이들 국가의 행정데이터를 사용한 연구는 해당 국가뿐만 아니라 한국 사회가 직면한 여러 사회경제적 문제의 해결이나 학계의 오랜 질문에 보편적으로 기여하는 경우가 많다. 가령 스웨덴의 행정데이터를 사용한 연구를 살펴보면, 복권 당첨 데이터와 건강보험 데이터베이스를 연계하여 복권 당첨으로 인해 예상하지 못한 방식으로 갑작스럽게 자산이 증가했을 때, 이러한 자산 증가가 본인의 건강과 자녀의 성장에 미치는 효과를 분석하거나(Cesarini et al., 2016), 노동시장에서 사고력에 대한 높은 보상이 실제 사고력의 발달로 이어지는지 살펴보는 연구가 있다(Hermo et al., 2022).

덴마크에서는 여성의 경력에 임신이 미치는 영향을 분석하거나(Lundborg et al., 2017), 소득세 변화가 국가 간 이민과 고소득자의 임금에 미치는 효과를 살펴보는 연구(Kleven et al., 2014)가 존재하였는데, 이러한 연구를 통해 행정데이터 활용의 유용성을 확인할 수 있다.

노동시장이나 보건의료 분야 이외에도, 시계열 자료를 활용하여 기업의 이동에 따른 도시성장과 산업구조의 변화를 관찰하는 연구(Imbert et al., 2022)와 같이 산업조직 분야에도 행정데이터가 활용되는 사례를 볼 수 있었다.

### 3. 세부 주제별 분포

경제학 분야 학술논문은 전미경제학회(American Economic Association)의 Journal of Economic Literature에서 정립한 JEL 분류 코드를 부여받는다. 논문의 게재가 확정되면 각 학술지의 편집위원회는 논문의 유형, 성격, 그리고 주제에 따라 논문의 저자가 제출한 관련 분야를 참고하여 논문에 분류 코드를 부여한다.

총 세 단계로 구성된 이 분류 기호는 알파뉴메릭 방식으로 표기하며, 알파벳으로 대분류를, 숫자로 중분류 및 소분류를 나타낸다. 예를 들어, L 기호는 산업조직 대분류를 나타내고, L1 기호는 시장구조, 기업전략 및 시장 성과를, L13 기호는 과점 및 기타 불완전시장을 다룬 논문임을 의미한다. 하나의 논문에 복수의 JEL 분류 기호가 부여되기도 한다.

이 분류 체계를 살펴보면 대분류를 기준으로 20가지로 구분되며(〈Table 1〉 참조), 소분류까지 고려하면 856가지 상세 분류로 학술논문의 주제 분야를 나눌 수 있다.

〈Table 1〉 JEL Categories

Code	Field	Code	Field
A	General Economics and Teaching	K	Law and Economics
B	History of Economic Thought, Methodology, and Heterodox Approaches	L	Industrial Organization
C	Mathematical and Quantitative Methods	M	Business Administration and Business Economics
D	Microeconomics	N	Economic History
E	Macroeconomics and Monetary Economics	O	Economic Development, Innovation, Technological Change, and Growth
F	International Economics	P	Political Economy and Comparative Economic Systems
G	Financial Economics	Q	Agricultural and Natural Resource Economics
H	Public Economics	R	Urban, Rural, Regional, Real Estate, and Transportation Economics
I	Health, Education, and Welfare	Y	Miscellaneous Categories
J	Labor and Demographic Economics	Z	Other Special Topics

앞서 논의한 것처럼 논문의 제목과 초록에 행정데이터 관련 주제어가 포함되었는지에 따라 논문을 두 그룹으로 분류한 후, 행정데이터를 사용한 논문의 분야별 분포와 행정데이터를 사용하지 않은 논문의 분야별 분포를 비교해보았다.

개별 논문에는 복수의 JEL 코드가 포함되기도 하는데, 한 논문이 복수의 대분류 기호를 포함(예: I21; J24) 하면, 각 대분류(예: I, J)에 해당하는 논문으로 간주하였다. 어떤 논문이 여러 분류 코드를 포함하고 있지만 대분류는 모두 같고 소분류 수준에서만 다른 경우(예: J13; J23; J31)는 대분류 J에만 속하는 것으로 처리하였다. 논문은 평균 약 2.6개의 다른 대분류 코드를 갖고 있었다(대분류가 하나만 있는 경우 17.4%, 대분류 코드가 2개 있는 경우 33.4%, 대분류 코드가 3개 있는 경우 28.7%, 대분류 코드가 4개 있는 논문은 15.9%, 대분류 코드가 5개 이상인 경우는

4.6%임).

본 연구에서는 이 분석을 통해 대략적으로나마 행정데이터를 사용한 연구의 비중이 높은 분야를 확인하고자 하였고, 관련 분석 결과를 <Table 2>에 제시하였다. 노동 및 인구 경제학 분야를 다루는 논문에는 대분류 코드 J가 부여되는데, 행정데이터를 사용한 것으로 분류된 109개의 논문 중 63.3%에 해당하는 논문에는 JEL 코드 J가 포함되어 있었다. 반면, 행정데이터를 사용하지 않은 것으로 분류된 논문 중 JEL 코드 J가 포함된 논문은 23.4%에 불과했다. 보건, 교육, 복지 분야에 속한 논문은 JEL 코드 I를 부여받는데, 행정데이터를 사용한 것으로 저자들이 분류한 109편의 논문 중 40.4%가 JEL 코드 I를 가지고 있었고, 이는 행정데이터를 사용하지 않은 것으로 분류한 논문의 경우인 15.3%와 비교할 때 여전히 큰 차이임을 볼 수 있다. 이와 유사한 경향은 재정학/공공경제학(H)과 법경제학(K) 분야에서도 확인된다.

<Table 2> JEL Distribution of Studies Using Administrative Data by Taxonomy

(Unit: %)

Category	Field	Administrative Data		Ratio (A/B)
		Used (A)	Not Used (B)	
J	Labor and Demographic Economics	63.3	23.4	2.7
I	Health, Education, and Welfare	40.4	15.3	2.6
D	Microeconomics	40.4	58.2	0.7
H	Public Economics	34.9	13.2	2.6
G	Financial Economics	19.3	19.7	1.0
O	Economic Development, Innovation, Technological Change, and Growth	17.4	18.6	0.9
E	Macroeconomics and Monetary Economics	11.9	19.2	0.6
K	Law and Economics	10.1	5.1	2.0
L	Industrial Organization	9.2	21.7	0.4
R	Urban, Rural, Regional, Real Estate, and Transportation Economics	9.2	7.5	1.2

경제학 이론 논문이 다수 포함되는 미시경제학 대분류 코드 D는 행정데이터를 사용하지 않은 그룹에서 더 높은 빈도로 관찰되었다. 행정데이터를 사용한 논문 중

19.3%의 연구는 재무경제학 분야를 의미하는 JEL 코드인 G가 포함되어 있는데, 행정데이터를 사용하지 않은 논문 중에서도 JEL 코드 G가 포함된 논문의 비중이 19.7%로 나타나 두 그룹 간 차이는 0.4%p에 불과했다.

이처럼 경제학의 여러 분야별로 행정데이터를 사용하는 연구의 비중은 다를 수밖에 없는데, 특히 노동 및 인구 경제학(J), 보건, 교육, 복지(I), 재정학/공공경제학(H), 법경제학(K) 분야의 연구에서 행정데이터를 사용하는 비중이 높음을 확인할 수 있었다.

#### IV. 학술 및 정책연구를 위한 국내 행정데이터 수요

한국은 이미 공공데이터 포털을 통해 다양한 분야의 데이터를 개방하려고 시도하고 있으며, 이러한 노력이 세계적으로 인정받아 OECD 국가 중 공공데이터 개방지수에서 3회 연속 1위를 차지하는 성과를 이루었다. 2020년부터 시행된 ‘데이터 3법’은 데이터 개방 및 공개를 더욱 활성화하기 위한 법적 기반을 마련하였다. 이 법은 데이터의 안전한 활용과 공유를 촉진하여 산업 발전과 사회 혁신을 도모하는 데 중요한 역할을 하고 있다. 그러나 이러한 적극적인 공공데이터 개방 움직임에도 불구하고 사회과학 분야에서 행정 데이터의 활용은 매우 저조하다.

한국 사회가 직면한 사회경제적 문제를 해결하기 위해 근거기반정책을 마련할 필요성에 대한 공감대는 이미 넓게 형성되어 있다. 이를 위하여 2023년 하반기에 한국경제학회는 교육부의 사회정책담당관실과 협업 가능성을 논의하였으며, 그 시작 단계로 경제학 내 여러 세부 분야에 걸쳐 행정데이터를 활용한 연구주제를 발굴하고 이에 필요한 행정데이터의 수요를 파악하는 작업을 진행하고자 학회원에게 연구계획서 제출을 의뢰하였다. 여기서는 이들이 제출한 12편의 연구계획을 간략히 요약하여 정리하면서 향후 학술연구와 정책연구를 위해 활용할 수 있는 행정데이터에 대한 수요를 제시하고, 행정데이터 간 연계의 필요성과 이러한 작업을 통해 가능한 연구의 잠재력과 정책적 기여의 가능성을 논의하고자 한다.

현대 사회가 직면한 다양한 사회경제적 문제들에 대처하기 위해서는 깊이 있는 증거 기반 연구가 필수적이다. 예를 들어, 코로나19가 교육에 미친 영향, 조기 치매의 사회경제적 영향, 일·가정 양립 정책이 출산율에 미친 효과, 주거 복지 정책의 효과 등은 장기적인 인과 효과 추정이 필요한 중요한 연구주제다. 과거에는 데

이터 접근성의 한계로 인해 관련 주제에 대한 연구가 어려웠지만, 행정데이터가 제공된다면 이러한 주제들에 관한 연구가 더욱 확대될 수 있다.

본 절에서는 한국경제학회 회원 11명이 제출한 행정 데이터를 활용한 12편의 연구계획서를 요약 정리한다. 이 분석을 통해 행정데이터 개방 시 진행할 수 있는 다양한 연구 사례를 가늠해보고, 개방 및 연계가 필요한 국내 행정데이터 수요를 살펴보고자 한다. 이는 행정 데이터 개방의 방향을 설정하는 데 중요한 기초 자료로 활용될 수 있다.

### 1. 한국의 행정데이터 활용 연구를 위한 해외 참고 사례

한국경제학회 회원이 제출한 연구주제를 검토하기에 앞서 제출된 연구계획서에 인용된 대표적인 해외 사례를 몇 가지 소개하고자 한다. 먼저 교육 분야 사례로 국내 연구자는 Cai et al. (2023)과 Jack et al. (2023)에 주목하고 있다. 이들은 학업 성취도 데이터를 활용하여 코로나19로 인한 등교 제한이 학업성취에 미치는 영향을 실증적으로 분석했다. 이 연구들은 등교 제한이 특히 저소득층 학생들에게 더 큰 부정적 영향을 미쳤다는 점을 밝혀냈다. 이러한 결과는 교육 정책 결정 과정에서 학생들 간의 불평등을 고려하는 데 중요한 역할을 하며, 모든 학생이 균등한 교육 기회를 누릴 수 있는 정책 설계에 이바지하고 있다.

세계 여러 나라에서는 개인의 생애 소득 및 가족·학력 정보를 다양한 행정데이터의 결합을 통해 구축하고 있으며, 해당 자료에 기반해 대학 교육이 평생 소득에 미친 영향과 국가의 고등교육 투자의 편익 계산에 관한 연구를 수행하고 있다. 노르웨이에서는 전 국민의 정보가 포함된 행정자료를 활용해 대학 교육이 생애 소득에 미친 영향을 분석하였다(Bhuller et al., 2017). 1년의 추가 교육은 평생소득을 11% 증가시키는 것으로 추정되어 대학 교육은 개인 측면에서 기대수익이 높은 투자임을 확인하였다. 유사한 연구는 영국에서도 진행되었는데, 특정 연령의 모든 개인에 대해 학력 정보와 과세자료를 연계해 영국의 대학 교육이 개인의 평생 소득에 끼친 영향 그리고 고등교육에서 정부지출의 편익을 계산하였다(Britton et al., 2020). 분석 결과를 보면 대학 교육으로 증가한 평생납부세금은 대학 교육에 필요한 비용을 크게 초과하는 것으로 추정되었으며, 대학 전공에 따라 재정투입의 장기 편익은 매우 이질적인 것으로 나타났다.

청소년 범죄 처벌이 향후 교육 수준, 취업률, 재범 위험에 미치는 인과 효과 추정은 경제학, 범죄학, 정책학의 중요 연구주제이며, 최근 국내의 주요 정책 이슈인 축법소년 연령 하향, 학교폭력예방법 개정 이슈 등과도 밀접하게 관련되어 있다. Aizer and Doyle (2015)은 법원 소년부에 송치된 청소년 범죄자에게 무작위로 판사 배정이 이루어진다는 점에 착안하여, 비슷한 유형의 범죄를 저질렀지만 각기 다른 성향의 판사에게 배정되었기 때문에 다른 처벌을 받은 소년범들의 재범률을 비교하였다. 이를 통해 저자들은 상대적으로 엄한 처벌을 받은 소년범의 고등학교 졸업률이 유의미하게 감소하며 성인이 된 이후의 범죄율이 유의미하게 증가함을 보였다. 이 연구에는 미국 시카고 지역의 교육청 학생 데이터(1990~2006), 소년부 개별 사건 데이터(1990~2006), 교정국 개별 재소자 데이터(1993~2008) 등이 결합된 행정데이터가 사용되었다. Bhuller et al. (2020)의 연구에서도 형사재판의 판사가 무작위로 배정된다는 점을 이용하여, 비슷한 범죄를 저질렀지만, 상대적으로 엄한 판사에게 배정되어 강한 처벌을 받은 피고인과 관대한 판사에게 배정되어 약한 처벌을 받은 피고인 간의 출소 후 취업률과 재범률 비교를 통해 징역형 기간의 증가가 출소 후 취업과 재범에 미치는 인과관계를 추정하였다. 이 연구에서는 노르웨이의 양형 데이터(2005~2014년)와 학력, 취업, 소득, 인구학적 특성 관련 행정데이터(1967~2016년)를 결합하여 처벌이 삶의 다양한 측면에 미치는 장기적 인과관계를 살펴보고 있다.

보건복지 분야에서는 고용보험 데이터를 활용한 다양한 연구들이 언급되었는데, 특히 육아휴직이 부모의 고용과 소득, 그리고 출산율에 미치는 영향을 다각도로 조명한 연구가 주목받고 있다. 예를 들어, Lalive and Zweimüller (2009)의 연구는 오스트리아 사회보장 데이터베이스(Austrian Social Security Database)를 활용하여 육아휴직 기간 확대가 출산율을 증가시키는 것을 보였다. 이 연구는 육아휴직이 단기적으로는 모의 고용과 소득을 감소시키지만, 장기적으로는 이러한 부정적 영향이 나타나지 않음을 확인하였다.

또한, Dahl et al. (2014)의 연구는 1967년부터 2006년까지 노르웨이 전 국민의 행정자료를 사용하여, 남성 근로자의 육아휴직 사용이 동료 남성들의 육아휴직 사용을 증가시킴을 보였다. 이는 육아휴직 정책이 남성의 육아 참여를 촉진하는 데 이바지할 수 있음을 시사한다.

더 나아가, 스웨덴(Angelov et al., 2016), 덴마크(Kleven et al., 2019), 핀란드

(Sieppi and Pehkonen, 2019), 오스트리아(Kleven et al., 2019) 등 여러 국가에서도 자녀 출산 및 육아가 부모의 노동시장 성과에 미치는 영향에 관한 연구가 수행되었다. 이러한 연구들은 육아휴직 정책이 가족과 노동시장에 미치는 복합적인 영향을 이해하는 데 중요한 기여를 하고 있다.

마지막으로 주거 복지 분야에 관한 연구에서는 국세청 소득신고 데이터를 활용한 Chetty의 연구가 주목받고 있다. Chetty et al. (2014)은 미국 국세청 자료를 활용하여, 부모와 자녀 소득 간의 관계를 살펴보고 지역별로 상관관계가 어떻게 나타나는지를 분석했다. 이 연구는 가구 형태, 소득, 대학 교육, 거주지 정보 등을 종합적으로 분석하여 세대 간 이동성의 지역별 차이를 드러냈다.

Chetty et al. (2016)은 1990년대에 미국에서 실시된 주거 정책을 활용해서 주택 바우처를 받은 그룹과 받지 못한 그룹 사이의 주거 환경의 차이가 자녀들의 교육 수준과 소득에 미치는 장기적 영향을 살펴봤다. 이 연구는 양질의 주거 환경이 자녀들의 장기적 성과에 긍정적 영향을 미칠 수 있음을 보여준다. 또한, Chetty and Hendren (2018)은 같은 데이터를 사용해서 어린 시절에 거주한 지역이 개인의 대학 진학률, 소득, 혼인율, 그리고 출산율에 미치는 영향을 살펴보고, 거주 지역이 개인의 삶의 질과 사회적 성과에 미치는 장기적인 영향을 규명하고 있다. 이러한 연구들은 주거 복지 정책이 단순한 주거 조건 개선을 넘어서 삶의 다양한 측면에 영향을 끼칠 수 있는 중요한 수단임을 보여준다.

〈Figure 3〉 Key Administrative Data Used in International Economics Research



〈Figure 3〉에는 해외 학술연구에서 자주 사용되어온 행정데이터 유형을 정리하

였다. 이러한 행정데이터를 활용한 연구는 정책 입안자들에게 효과적인 정책 결정을 위한 데이터 기반 근거를 제공하는 중요한 역할을 하고 있다. 한국 또한 해외와 마찬가지로 학업성취도 데이터, 고용보험 데이터, 국세청 소득신고 데이터와 같은 다양한 행정 데이터를 보유하고 있다. 향후 이러한 행정데이터의 개방은 한국의 구체적인 상황과 요구에 맞는 맞춤형 정책연구를 가능하게 할 것이며, 이를 통해 더욱 정밀하고 효과적인 정책 방향을 찾아가는데 기여할 수 있다.

## 2. 행정데이터 활용 연구계획 소개

한국경제학회 회원 11명에게 의뢰하여 제출받은 국내 행정 데이터 활용 연구계획 12편의 주제는 <Table 3>에 제시하였다. 이들 연구계획서에서 다루는 주제는 노동, 교육, 보건, 재정, 그리고 산업조직론 분야에 걸쳐 다양하게 분포하고 있다.

연구계획서의 내용을 보면 현재 국내 연구자들이 활용할 수 있는 공개된 자료로는 분석할 수 없고 행정데이터의 활용이 필수적임을 알 수 있었다. 예를 들어, 교육 분야의 경우 해당 주제의 분석을 위해서는 장기간에 걸쳐 대표성을 담보할 수 있는 성취도 자료가 필요하거나, 법무부 자료와 교육부 자료의 개인 수준 연계가 필요한 사례가 제안되었다.

보건 분야의 경우에는 치매 발병이나 외국인 건강보험 적용 사례처럼 기존의 서베이 자료만으로는 분석 표본을 확보하기가 어려워 관련 연구가 매우 제한적인 주제가 제출되었다. 육아휴직 정책이나 일·가정 양립 관련 정책 연구의 경우에도 행정데이터 간 연계를 통해 자료가 확보되어야 다양한 하위 집단을 심층적으로 살펴보고 정책의 실질적 효과에 대한 분석이 가능하므로 행정데이터의 활용이 절대적으로 필요한 상황임을 확인할 수 있었다.

주거 복지와 관련한 두 연구에서는 공통으로 공공임대주택 입주자 선정 과정의 무작위추첨 방식의 활용이 제안되었다. 무작위추첨을 통해 선정되지 못한 이들이 정책 효과 분석에 필요한 적절한 비교군으로 역할을 할 수 있기 때문에, 이를 통해 관련 정책의 인과 효과를 식별하고 그 결과를 정책에 효과적으로 활용할 수 있다. 하지만, 아직 이러한 데이터가 학술연구나 정책연구에는 활용되지 못하고 있다.

소상공인에 대한 연구계획을 보면 이미 중소벤처기업부가 DB로 구축하여 내부적으로 활용하고 있는 자료를 이용하면 분석을 수행할 수 있다. 하지만, 이미 분석



가능한 수준으로 정제된 이러한 소중한 자료조차 연구 목적으로 학계에 공유되지 못하고 있는 현실을 개선하기 위해 학계와 정부의 적극적인 협력이 필요하다.

〈Table 3〉 List of Research Plans Using Administrative Data in South Korea

No.	Title	Data Demand
1	고교평준화가 학생들의 학업성취도, 학습경험, 웰빙에 미친 영향	건강보험DB, 국가수준학업성취도평가
2	고등교육 재정부자의 장기편익에 관한 연구	출신 고교 및 대학교 정보, 수학능력시험, 국세청DB
3	코로나로 인해 발생한 등교 일수 감소가 청소년의 학업성취도, 비인지 기능 및 학업 불평등에 미치는 영향 분석	국가수준학업성취도평가, 수학능력시험
4	학군이 교육성과에 미치는 영향	초중등 교육 정보, 건강보험DB, 국세청DB
5	행정 데이터를 이용한 청소년 범죄 처벌의 장기적 인과 효과 분석	초중등 교육 이력, 국세청DB, 법무부DB(청소년 범죄 관련)
6	(조기)치매 발병이 본인과 가족의 노동 성과 및 삶에 미치는 영향	건강보험DB, 고용보험DB, 국세청DB
7	국내 유입인구 지역가입자 당연적용제도의 효과 분석	건강보험DB, 법무부 출입국정보
8	육아휴직 정책이 근로자의 경제활동 및 출산에 미치는 영향	건강보험DB, 고용보험DB
9	일·가정 양립가능성과 결정요인에 관한 연구	건강보험DB, 고용보험DB, 교육부 개인 학력 정보
10	주거복지정책의 효과 분석	공공임대주택 청약·당첨 정보, 건강보험DB, 학업성취 및 성과
11	주거지원정책이 자녀 교육에 미치는 영향: 공공임대주택 무작위추첨 행정데이터를 이용한 실증분석	공공임대주택 청약·당첨 정보, 국가수준학업성취도평가, 수능
12	소상공인에 대한 정부 보조와 대출 정책의 효과 및 효율성 분석	중소기업 지원 사업 통합 관리 시스템(SIMS) DB

〈Table 3〉에 정리한 것처럼 제출된 연구계획서에는 고용보험 및 건강보험 데이터, 수학능력시험 및 국가수준학업성취도평가 자료, 국세청 소득자료 및 공공임대주택 청약 자료가 반복적으로 등장한다. 다수의 연구자는 이들 자료의 연계를 통한 학술연구와 정책연구의 잠재력을 높이 평가하고 있으며, 이러한 자료가 학술연구에

조속히 활용되기를 바라는 높은 관심과 활용 의지를 보여주었다. 특히 고용보험 및 건강보험 데이터와 국제청 소득자료의 활용과 연계는 여러 연구주제에서 필수적인 행정 데이터로 요구되고 있었다. 이들 자료에 대한 접근이 가능해진다면 향후 다양한 연구 분야에서 폭넓게 활용될 수 있을 것으로 보인다.

## V. 맺음말

본 연구에서는 먼저 행정데이터를 활용한 경제학 분야 해외 학술연구를 검토하였다. 다음으로는 한국경제학회 회원이 제출한 연구계획서를 토대로 국내 행정데이터를 활용해서 수행할 수 있는 다양한 연구를 소개하고, 연구에 필요한 관련 행정데이터 수요를 정리하였다. 본 논문에 담긴 내용은 향후 정부가 증거기반 정책을 위한 행정데이터 공개 대상과 행정데이터 간 연계 작업의 우선순위를 결정하는 과정에서 유용하게 활용될 수 있을 것이다. 아울러 본 연구를 통해 제안된 연구자들의 행정데이터 수요가 반영되고 행정데이터 접근성이 개선된다면, 정책의제 개발과 정책평가를 수행하는 실증연구가 보다 활발하게 진행될 수 있을 것으로 기대한다.

앞서 논의한 것처럼 증거기반 정책에 관한 관심과 이를 위해 행정자료를 활용하려는 노력은 세계적인 흐름이다. 해외 학계에서는 이러한 흐름을 반영하여, 행정데이터를 활용해서 정책연구를 수행하기 위한 사례와 방법론을 소개하고, 관련한 법적 이슈까지 논의하는 안내서가 발간되기도 하였다(Cole et al., 2020). 하지만 국내에서는 증거기반 정책연구를 위한 행정자료의 활용에 대한 논의가 과거부터 진행됐음에도, 아직 해외의 사례에 비추어 미흡한 점이 많다(김재진 외, 2020; 유종성 외, 2020; 이영욱 외, 2020).

국내 행정데이터를 활용한 학술 및 정책연구를 활성화하기 위해서는 개방할 수 있는 데이터를 확대함과 동시에 해당 데이터들을 제공하는 방식이나 범위가 일관되게 유지될 필요가 있다. 가령 교육 분야의 사례를 보면, 한때는 수학능력시험이나 국가수준학업성취도평가 개인별 전수자료가 학교명과 함께 제공되었다. 국내외 연구자들은 이 자료를 활용하여 국제적으로 경쟁력 있는 연구를 수행하고 해외 우수 학술지에 연구 결과가 게재되었다. 하지만 2010년 중반 이후로는 에듀데이터서비스(EDSS)를 통해 전수자료가 아닌 표본 추출된 자료만 제공되고 있고, 학교명이나 지역 정보 등을 식별자로 사용해서 연구자가 다른 데이터를 연결해서 분석할 수 있

는 기회가 제한되면서 해당 자료를 사용하려는 연구자의 관심은 오히려 줄어들었다.

다음으로 개별적으로 사용되던 행정데이터 간 연계를 통해 활용 가치를 높일 필요가 있다. 고용보험DB나 건강보험DB는 개별 데이터만으로도 매우 중요한 정보를 담고 있고, 이러한 자료에 접근할 수 있는 기회가 부분적으로나마 확대되면서 다양한 연구에 활용되고 있다. 하지만, 이 두 자료의 정보를 개인 수준에서 결합하여 분석할 수 있다면 학술적으로 그리고 정책적으로 중요한 연구들이 가능해지고 확대될 수 있을 것이다. 장기적으로는 이들 자료가 학생과 교사에 대한 풍부한 정보를 보유한 교육행정정보시스템(NEIS) DB나 소득과 자산에 대한 정보를 제공할 수 있는 국세청 행정자료와 결합되어야 한다. 이들 자료를 연구에 활용할 수 있게 된다면, 행정데이터를 분석한 해외의 영향력 있는 연구 사례처럼 한국 사회가 직면한 사회 문제 해결을 위해 이를 활용할 수 있을 것이며, 해외 연구에서 한층 더 나아간 선도 사례 또한 우리가 제시할 수 있는 기회도 열릴 것이다.

행정데이터는 정확성과 규모 측면에서는 매우 강점이 있지만, 개인이나 가구에 대한 인구 사회학적 정보는 부족한 경우가 많다. 이를 보완하기 위해서는 이미 수집된 다양한 서베이 자료에 행정데이터를 통해 확보된 정보를 연결하는 방식으로 각 데이터의 강점을 살리고 다양한 연구의 가능성을 확보할 수도 있다.

다음으로는 행정데이터 활용의 관점에서 물리적 제약을 언급하고자 한다. 통계청이나 정부 기관에서 행정DB를 연계하여 좋은 자료를 구축하더라도 이를 연구자가 접근하기 어렵다면 활용은 제한될 수밖에 없다. 현재 통계청, 건강보험공단, 그리고 서울시가 운영하는 데이터분석센터가 존재하지만, 연구자 입장에서는 평일 근무 시간에 해당 장소에 방문해서 장기간 연속적으로 데이터 분석을 수행하기는 어렵다. 보안이 매우 중요한 자료라면 외부의 원격접근이 차단된 분석 환경을 제공해야 하는데, 그런 경우에는 분석 센터를 확대하는 방향으로 물리적인 접근성을 높이고, 연속적이고 집중적인 분석이 가능하도록 분석 센터의 운영 시간도 현행 평일 공공기관 근무시간보다는 확대 운영될 필요가 있다. 장기적으로는 원격접근을 통해 분석할 수 있는 시스템을 보다 확대할 필요가 있다. 아울러 연구가 학술지에 게재되기까지는 길게는 2-3년 이상 소요되기도 한다는 점도 고려될 필요가 있다. 연구 수행 과정에서 수정 및 추가 분석 작업이 반복적으로 필요한데, 물리적인 접근성 제약뿐만 아니라 데이터 접근을 위한 반복적인 행정 절차와 비용 부담은 행정데이터

활용에 큰 장애요인이 된다.

잘 알려진 바와 같이 행정데이터 간 연계의 정확성을 높이고, 행정데이터와 서베이 자료와의 연계를 위해서는 주민등록번호를 연계키로 사용할 필요가 있지만 현행 법제도 하에서는 어려움이 많다. 향후 이와 관련된 법적 문제 및 개인정보 보호를 위한 비식별화 처리 방식을 고민하고 지원할 필요가 있다.

아울러 행정데이터의 활용 가치를 높이기 위해서는 정부와 학계의 협력을 통해 과제를 발굴하고 공동 연구를 수행할 필요가 있다. 한국의 행정데이터 수준과 연구자의 분석 역량은 다른 선진국과 비교해도 매우 뛰어나다. 이러한 우수한 자원을 적극적으로 활용할 수 있다면 향후 국내에서도 사회 문제 해결과 효과적 정책 수립 및 평가를 위한 근거 기반 정책이 확대될 수 있을 것으로 기대한다.

무엇보다 행정데이터의 공개를 통해 여러 정부 기관이 담당하고 있는 다양한 사회 현안에 대한 진단과 엄밀한 평가가 가능해질 것이며, 이러한 결과를 바탕으로 정책 대안을 발굴할 수도 있을 것으로 기대한다. 아울러 데이터를 외부에 개방하고 제공하는 과정에서 데이터의 오류가 발견되고 수정되면서 정부가 보유한 자료의 정확성이 높아질 수 있고, 데이터 관리 및 운영 방식에 대해 이용자에게 제안받고 소통하는 과정을 통해 데이터 관리 시스템의 효율성도 한층 개선될 수 있다.

현재 한국경제학회 회원이 참여하고 있는 연구가 개인정보보호위원회의 데이터 선도 사례로 선정되어 사회보장 정보시스템 행복e음과 건강보험DB 연계 작업이 진행되고 있다. 또한 여러 회원이 건강보험공단 빅데이터 연구 전문위원으로 활동하면서 사회경제적 건강 취약 집단을 발굴하기 위한 연구를 진행하는 등 행정데이터를 학술 연구에 활용하기 위해 노력을 기울이고 있다. 이러한 연구자와 기관의 노력을 정부가 더 적극적으로 지원하고, 공동 작업이 더욱 확대될 수 있는 기회를 제공해주었으면 한다. 아울러 행정데이터를 활용한 연구 성과를 공유할 수 있는 기회가 증가하고, 경제학뿐만 아니라 여러 학문 분야에서 활동하는 연구자들이 함께 모여서 다양한 관점에서 공동 연구를 수행하고 생산적인 논의를 이어갈 수 있는 교류의 장도 마련되기를 바란다.

## ■ 참 고 문 헌

1. 김재진 · 노대명 · 이현주 · 정찬미, “증거기반 정책 인프라 확충을 위한 소득 관련 행정데이터 연계 · 통합방안 연구” 한국조세재정연구원, 2020.  
(Translated in English) Kim, Jaejin, Daemyung Noh, Hyunjoo Lee, and Chanmi Jung, “Study on How to Link and Integrate Income-related Administrative Data to Expand Evidence-based Policy Infrastructure,” Korea Institute of Public Finance, 2020.
2. 유종성 · 전병유 · 신광영 · 이도훈 · 최성수, “증거기반 정책연구를 위한 행정자료의 활용,” 『한국사회정책』, 제27권 제1호, 2020, pp.5-37.  
(Translated in English) You, Jong-Sung, ByungYou Cheon, Kwang-Yeong Shin, Dohoon Lee, and Seongsu Choi, “Using Administrative Data for Evidence-Based Policy Research,” *Korea Social Policy Review*, Vol. 27, No. 1, 2020, pp.5-37.
3. 이영욱 외, “재정성과 향상을 위한 증거기반 정책평가,” 한국개발연구원, 2020.  
(Translated in English) Lee, Youngwook, et al., “Evaluating Evidence-based Policies to Improve Financial Performance,” KDI, 2020.
4. 최재성, “빅데이터 시대의 사회과학 연구,” 전환과 변동의 시대 사회과학, 최훈석(편), 성균관 대학교출판부, 2022.  
(Translated in English) Choi, Jaesung, “Social Science Research in the Age of Big Data,” In Hoonseok Choi (Eds.), *Social Sciences in an Era of Transition and Change*, Sungkyunkwan University Press, 2022.
5. 황윤재 · 최재성, “사회분야 데이터 기반 정책 연구의 현황과 과제,” 교육부, 2023.  
(Translated in English) Hwang, Yoonjae, and Jaesung Choi, “Status and Challenges of Data-driven Policy Research in Social Science,” Ministry of Education, 2023.
6. Aizer, Anna, and Joseph J. Doyle Jr., “Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges,” *The Quarterly Journal of Economics*, Vol. 130, No. 2, 2015, pp.759-803.
7. Angelov, N., P. Johansson, and E. Lindahl, “Parenthood and the Gender Gap in Pay,” *Journal of Labor Economics*, Vol. 34, No. 3, 2016, pp.545-579.
8. Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad, “Incarceration, Recidivism, and Employment,” *Journal of Political Economy*, Vol. 128, No. 4, 2020, pp.1269-1324.
9. Bhuller, M., M. Mogstad, and K. G. Salvanes, “Life-cycle Earnings, Education Premiums, and Internal Rates of Return,” *Journal of Labor Economics*, Vol. 35, No. 4, 2017, pp.993-1030.
10. Britton, J., D. Lorraine, V. Laura, and B. Waltmann, “The Impact of Undergraduate Degrees on Lifetime Earnings,” Department for Education Report, 2020.
11. Cai, X., J. Fu, M. Luan, and X. Tang, “Assessing Inequality in the School Closure Response to COVID-19,” *China Economic Review*, Vol. 80, 2023.
12. Cesarini, D., E. Lindqvist, R. Östling, and B. Wallace, “Wealth, Health, and Child Development: Evidence from Administrative Data on Swedish Lottery Players,” *The*

- Quarterly Journal of Economics*, Vol. 131, No. 2, 2016, pp.687-738.
13. Chetty, R., N. Hendren, P. Kline, and E. Saez, "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States," *The Quarterly Journal of Economics*, Vol. 129, No. 4, 2014, pp.1553-1623.
  14. Chetty, R., and N. Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *The Quarterly Journal of Economics*, Vol. 133, No. 3, 2018, pp.1107-1162.
  18. Chetty, R., N. Hendren, and L. F. Katz, "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, Vol. 106, No. 4, 2016, pp.855-902.
  16. Chetty, R., D. J. Deming, and J. N. Friedman, "Diversifying Society's Leaders? The Causal Effects of Admission to Highly Selective Private Colleges," National Bureau of Economic Research Working Paper No. w31492, 2023.
  17. Cole, S., I. Dhaliwal, A. Sautmann, and L. Vilhuber, Handbook on using Administrative Data for Research and Evidence-based Policy, 2020. <https://admindatahandbook.mit.edu/book/v1.0-rc5/index.html>.
  18. Dahl, G. B., K. V. Løken, and M. Mogstad, "Peer Effects in Program Participation," *American Economic Review*, Vol. 104, No. 7, 2014, pp.2049-2074.
  19. Dynarski, S., C. J. Libassi, K. Micheltore, and S. Owen, "Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-income Students," *American Economic Review*, Vol. 111, No. 6, 2021, pp.1721-1756.
  20. Einav, L., J. and Levin, "Economics in the Age of Big Data," *Science*, Vol. 346, 2014.
  21. Herno, S., M. Päällysaho, D. Seim, and J. M. Shapiro, "Labor Market Returns and the Evolution of Cognitive Skills: Theory and Evidence," *The Quarterly Journal of Economics*, Vol. 137, No. 4, 2022, pp.2309-2361.
  22. Imbert, C., M. Seror, Y. Zhang, and Y. Zylberberg, "Migrants and Firms: Evidence from China," *American Economic Review*, Vol. 112, No. 6, 2022, pp.1885-1914.
  23. Jack, R., C. Halloran, J. Okun, and E. Oster, "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts," *American Economic Review: Insights*, Vol. 5, No. 2, 2023, pp.173-190.
  24. Kirkeboen, L. J., E. Leuven, and M. Mogstad, "Field of Study, Earnings, and Self-selection," *The Quarterly Journal of Economics*, Vol. 131, No. 3, 2016, pp.1057-1111.
  25. Kleven, H. J., C. Landais, E. Saez, and E. Schultz, "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark," *The Quarterly Journal of Economics*, Vol. 129, No. 1, 2014, pp.333 - 378.
  26. Kleven, H., C. Landais, and J. Sogaard, "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, Vol. 11, No. 4, 2019, pp.181-209.
  27. Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller, "Child Penalties across Countries: Evidence and Explanations," *AEA Papers and Proceedings*, Vol. 109,

- 2019, pp.122-126.
28. Kopczuk, W., E. Saez, and J. Song, "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937," *The Quarterly Journal of Economics*, Vol. 125, No. 1, 2010, pp.91-128.
  29. Lalive, R., and J. Zweimüller, "How does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments," *The Quarterly Journal of Economics*, Vol. 124, No. 3, 2009, pp.1363-1402.
  30. Lundborg, P., E. Plug, and A. W. Rasmussen, "Can Women Have Children and a Career? IV Evidence from IVF Treatments," *American Economic Review*, Vol. 107, No. 6, 2017, pp.1611-1637.
  31. Piketty, T., and E. Saez, "Inequality in the Long Run," *Science*, Vol. 344, 2014, pp.838-843.
  32. Sieppi, A., and J. Pehkonen, "Parenthood and Gender Inequality: Population-based Evidence on the Child Penalty in Finland," *Economics Letters*, Vol. 182, 2019, pp.5-9.

## Current Status and Challenges of Data-Based Research in Economics\*

Jaesung Choi\*\* · Yoon-Jae Whang\*\*\*

### Abstract

Interest in evidence-based policies and efforts to use administrative data for this purpose are global trends, and research that links administrative data held by the government and uses it to identify causal relationships plays an important role in policy development and evaluation. In Korea, discussions on the use of administrative data for evidence-based policy research have been steadily conducted in the past, but the results are still very limited. This study aims to identify the potential for domestic research using administrative data and lay the groundwork for future cooperation between the government and academia. To achieve this objective, we collected bibliographic information and JEL classification codes for articles published in four leading journals in the field of economics and used them to examine trends in overseas research using administrative data. Furthermore, based on 12 research plans submitted by members of the Korean Economic Association, we introduced research that can be conducted using domestic administrative data and summarized the administrative data required for such research. This study will be useful for the government in prioritizing the targets of administrative data disclosure for evidence-based policies and in determining the priority of linkage between various administrative data. Additionally, if the demand for administrative data proposed in this study is reflected and accessibility is improved, it is expected that empirical research on policy agenda development and policy evaluation will rapidly expand in South Korea.

**Key Words:** administrative data, evidence-based policy

**JEL Classification:** A1

---

*Received: Feb. 22, 2024. Revised: March 13, 2024. Accepted: March 20, 2024.*

\* This paper is a revised manuscript of the authors' final report 'Status and Challenges of Data-driven Policy Research in Social Science', submitted to the Ministry of Education in 2023. The authors would like to thank Hyerin Seo, PhD student, Sungkyunkwan University, and Dohoon Lee, MA student, Sungkyunkwan University, for their help in surveying the literature and analyzing data during the writing of this paper. We are also grateful to the members of the Korea Economic Association who prepared the research plan at short notice to expand the possibilities and horizons of research on the use of administrative data in South Korea. We hope that this effort will pave the way for data-driven policy research in South Korea and expand cooperation between the government and academia in the future.

\*\* First Author, Associate Professor, Department of Global Economics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea, Phone: +82-2-760-0146, e-mail: jaesungc@skku.edu

\*\*\* Corresponding Author, Professor, Department of Economics, Seoul National University, Bld. 16, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea, Phone: +82-2-880-6362, e-mail: whang@snu.ac.kr