북한경제 연구에서 대규모 행정데이터의 활용 가능성

이종민 | 한국은행 부연구위원 | ljm1208@bok.or.kr

Ⅰ. 머리말

북한경제를 탐구하는 연구자에게 가장 큰 어려움은 무엇일까? 여러가지 대답이 있을 수 있겠지만 '신뢰성 있는 데이터의 부족'이라는 대답은 빼놓을 수 없을 것이다. 일반적인 경제학 분야에서는 아주 구체적인 한 가지 사실을 밝혀내기 위해서도 매우 풍부한 실증적 근거가 제시되는 경우가 많다. 같은 주제를 두고 개별 연구진들이 서로 다른 데이터를 이용하여 분석을 수행하고 그로부터 얻은 연구 결과들이 쌓여 가면서 해당 주제에 대한 이해의 폭을 넓혀간다. 그러나 북한경제는 연구해야 할 주제에 비해 연구의 재료가 될만한 데이터가 양적, 질적으로 매우 부족한 상황이다. 충분하지 않은 실증적 근거를 가지고 광범위한 영역을 다루어야 하다 보니 연구자의 주관적 해석이 실증적 근거보다 앞서는 것 같은 느낌을 받을 때도 있다. 특히 아직 '내공'이 부족한 초보 연구자 입장에서는 신뢰할 만한 자료가 부족한 부문에 대해 '전문가'로서의 의견을 요구받을 때면 퍽 난감해지곤 한다.

이런 점에서, 최근 북한경제 연구의 각 부문에서 새로운 자료원에 대한 접근을 통해 돌파구를 마련하고자 하는 시도가 활발해진 것은 매우 고무적인 일이다. 김규철(2020), Kim(2021), Kim *et al.*(2022), 김다울 외(2022)는 모두 인공위성에서 추출된 데이터를 활용하여 북한 내부의 전반적 경제활동 및 후생수준 추이, 지역별, 산업별 동향, 식생 현황 등을 파악하고자 하였다. 김수현・손욱(2020), 이가영(2021), 이해정 외(2021)는 북한의 원전을 텍스트마이닝 기법으로 분석하여 시기별 북한 경제정책의 특징적인 모습을 제시하였다. 또한 산업연구원에서는 그동안 북한 내부 언론보도 자료를 이용하여 축적해 둔 북한의 기업 정보를 DB화하여

여러 연구에 활용하고 있다(이석기 외, 2015).

본고는 이러한 움직임의 연결선상에서, 북한경제 연구에서 그동안 많이 주목받지 못하였던 '행정 빅데이터' 활용의 가능성에 대해 논하고자 한다.

Ⅱ. 행정데이터의 정의, 사례, 성격

행정데이터(administrative data)는 공공기관 등이 행정업무를 수행하는 과정에서 수집하고 기록하는 데이터이다. 행정안전부의 주민등록데이터, 법무부의 내·외국인 출입국데이터, 한국은행의 가계부채 DB, 건강보험공단의 국민건강정보 DB 등이 대표적인 행정데이터이다. 이러한 자료들이 경제학 분석에 활용되기 위해서는 정보의 손실이 최소화될수록, 즉 최대한 비집계 수준(disaggregate level)에서 제공될수록 좋다. 그러나 아쉽게도 기관들이 보유하고 있는 정보의 민감성, 정보 공유 플랫폼 부재 등으로 인해 집계된 수준에서만 제공되는 경우가 많다. 예를 들어 법무부는 외국인의 입국과 관련하여 여권의 기본 정보 및 입국 및 비자발급 시 기재한 정보, 생체정보 등을 수집하고 있으나, 일반 대중에게 공개할 수 있는 데이터는 국적-항구별, 국적-체류자격별로 집계된 출입국 인원 통계뿐이다.[1] 북한경제 연구에서 많이 활용되는 무역자료도 행정데이터 중 하나이지만 일반적으로 연구자들이 접근 가능한 것은 품목별로 집계된 데이터여서 심화된 계량분석을 진행하는 데에는 한계가 존재한다.

행정데이터는 수집 과정과 목적 면에서 기존 연구에서 주로 사용되는 자료들, 특히 설문조사 자료와 큰 차이가 있다. 설문조사 자료는 특정 주제의 조사·연구를 목적으로 수집되며, 조사 대상자의 응답을 바탕으로 작성된다. 반면, 행정데이터는 행정 업무를 수행하는 과정에서 자연스럽게 수집되고 행정 주체의 관측과 기록으로 작성된다. 따라서 응답자의 기억과 주관에 의존하는 서베이 자료에 비해 오류의 가능성이 작고 정확도가 높은 것으로 평가된다. 특히 대규모 행정데이터를 비집계 수준에서 접근할 수 있다면 모집단에 가까운, 심지어는 모집단 전수자료를 이용한 연구도 가능하며, 원자료를 연구의 목적에 따라 다양한 방식으로 재가공하여 활용할 수 있다. 즉, 경제학 실증연구에 있어서 대규모 행정데이터는 자료의 정확성과 정보의 양, 표본의 크기 등 여러 측면에서 이점을 갖는다고 할 수 있다.

연구 활용도 면에서 행정데이터의 한계도 존재한다. 우선 애초에 조사·연구를 목적으로

1) 최서리 외, 『법이민분야 공공데이터 현황과빅데이터 구축과제: 법무부 데이터와 고용·교육 데이터 연계를 중심으로』, 이민정책연구원 정책브리프시리즈 2021-01; 법무부, 내외국인 국적 및 항구별 출입국 현황(https://www.data.go.kr/data/3074947/fileData.do, 검색일: 2022. 12. 12).

수집된 자료가 아니기 때문에, 연구자가 관심 있는 정보들이 충분히 존재할 것이라고 기대하기 어렵다. 또한 하나의 연구를 수행하는 데 필요한 자료들을 업무의 성격에 따라 개별 기관에서 산발적으로 보유하고 있는 경우도 많다. 서베이 자료와 달리 추가적인 조사를 통한 데이터의 개선이 불가능하다는 것도 단점이다.

Ⅲ. 북한경제 연구에의 활용 가능성

앞서 언급한 행정데이터 안에는 국내에 입국해 있는 북한이탈주민에 대한 정보도 당연히 포함되어 있다. 아직 초기 단계이기는 하지만 대규모 행정데이터를 활용한 연구들도 존재한다. 정승호 외(2018)는 NICE평가정보가 보유하고 있는 개인 신용정보 데이터를 이용하여 북한이탈주민의 신용거래 행태에 대해 분석하였고, Pak *et al.*(2011), 위혜승(2018)은 각각 한국 입국 시와 건강검진 수검 시 측정된 북한이탈주민의 신장 자료를 이용하여 출생 시기별 북한 내부의 생활수준을 평가하였다.

이렇게 행정데이터가 북한경제 연구에 활용되려면 무엇보다 해당 데이터베이스에 북한이탈주민을 일반 남한주민으로부터 식별할 수 있도록 해주는 정보가 포함되어 있어야 한다. 안타깝게도 대부분의 경우에는 이것이 공공기관의 행정업무 수행과 관련이 없는 정보이므로 북한이탈주민 식별이 어렵다. 예외적으로 최근 광범위한 접근이 가능해진 건강보험공단의 국민건강정보 DB는 건강보험 수급자격데이터를 바탕으로 북한이탈주민을 쉽게 식별할 수 있어 북한경제 연구에의 활용 가능성이 높을 것으로 기대된다.

국민건강정보 DB는 2002년부터 2020년까지 국민건강보험자격자(전 국민)의 의료 이용 정보, 건강검진 정보 등을 데이터베이스화하여 제공하고 있으며, 매년 말 직전 해의 자료가 업데이트된다. 또한 일반 대중에 공개되어 있는 표본데이터 외에도 연구의 목적에 따라 맞춤형 DB를 신청할 수 있게 되어 있어 표본 재구성 및 주제별 세부정보를 포함한 데이터베이스 접근이 가능하다.[2] 해당 DB에서 제공하는 의료 이용 관련 정보와 건강검진 자료를 이용하면 주요 질병의 유병 여부 및 건강 관련 각종 측정지표 등을 확인할 수 있으며, 가입자별 분류를 통해 노동시장 참여 여부도 파악이 가능하다. 이를 북한이탈주민 대상 연구에 활용한다면 인적자본 수준, 남한 사회에 대한 적응 문제 등을 다룰 수 있어 분석의 가치가 높다. 또한 영유아기의 성장환경을 반영하는 지표이기 때문에 출생연도에 따라 분석하면 북한 내부

2) 국민건강보험자료 공유서비스(https://nhiss.nhis.or.kr/bd/ay/bdaya001iv.do, 검색일: 2022. 12. 13).

보건·영양 환경의 장기적 추세를 가늠하는 데에도 활용할 수 있다.

이와 같은 대규모 행정데이터의 활용은 북한경제 연구의 발전에 기여할 수 있는 잠재력을 가지고 있는 것으로 보인다. 특히 기존 북한이탈주민 설문조사의 문제점 중 하나로 지적되어 온 낮은 응답률, 기억 오류 등에 의한 편의(bias) 발생 가능성을 크게 낮출 수 있다. 또한 전수에 가까운 표본 추출이 가능하므로 북한이탈주민의 남한 사회 정착 등과 관련하여 정책적 함의가 높은 연구를 수행할 수 있다. 다만, 다양한 부문의 북한 내부동향을 파악하는 데에는 한계가 존재한다. 이는 북한 거주 당시 경험에 해당하는 정보를 남한의 행정데이터를 통해 파악하기 어렵기 때문이다. 따라서 기존의 설문조사를 통한 북한이탈주민 연구와 상호 보완적으로 연계될 수 있다면 가장 바람직할 것이다. 또한 북한이탈주민의 구성과 성격이 전체 북한주민들과 다를 수 있다는 점은 행정데이터를 이용한 연구에서도 동일하게 유념해야 할 부분이다. 앞으로 대규모 행정데이터를 활용한 다양한 연구들이 북한경제에 대한 이해를 풍성하게 하는 데 널리 활용될 수 있기를 기대해 본다.

참고문헌

김규철, 『새로운 데이터로 추정한 북한의 소득과 후생의 장기 추세: 1인당 GDP 추정을 중심으로』, KDI 정책연구시리즈 2020-04, 세종: 한국개발연구원, 2020.

김다울·김범환·한하린·이대은, 『북한 기후변화 적응을 위한 국제협력방안: 농업과 자연재해를 중심으로』, KIEP 전략지역심층연구, 2022(근간예정).

김수현 · 손욱, 「북한 「경제연구」로 분석한 경제정책 변화: 텍스트마이닝 접근법」, BOK경제연구 2020-6호, 2020.

위혜승, 「기근이 신체에 미친 효과: 북한이탈주민의 키를 통해 살펴본 코호트별 분석」, 『보건경제와 정책연구』, 제24권, 제4호, 2018.

이가영, 「텍스트마이닝을 통한 김정은 시기 경제정책 연구: 토픽 모델링 분석을 중심으로」, 『북한학연구』, 제17권, 제2호, 2021.

이석기 · 심완섭 · 이승엽 · 빙현지 · 김창모, 『북한 공식매체를 통해 본 산업정책 및 주요 산업 · 기업 변화 실태』, KIET 정책자료 2015-260, 세종: 산업연구원, 2015.

이해정 · 신훈식 · 이혜진, 「텍스트 마이닝을 통한 김정은 시대 경제정책 분석」, 『국가전략』, 제27권, 제1호, 2021.

정승호 · 민병기 · 김주원, 「북한이탈주민의 신용행태에 관한 연구」, BOK경제연구 2018-11호, 2018.

최서리·유민이·김영근, 『법이민분야 공공데이터 현황과 빅데이터 구축과제: 법무부 데이터와 고용·교육 데이터 연계를 중심으로』, 이민정책연구원 정책브리프시리즈 2021-01, 서울: 이민정책연구원, 2021.

Kim, Dawool, "Determinants of Regional Economic Performance in North Korea: Evidence from Satellite Nighttime Lights," 서울대학교 박사논문, 2021.

Kim, Jihee, Kyoochul Kim, Sangyoon Park, and Chamg Sun, "The Economic Costs of Trade Sanctions: Evidence from North Korea," Available at SSRN: https://ssrn.com/abstract=4032573, 2022.

Pak, Sunyoung, Daniel Schwekendiek, and Hee-Kyoung Kim, "Height and Living Standards in North Korea, 1930s-1980s," *Economic History Review*, vol. 64, Issue S1, 2011.