

북한이탈주민 설문조사 자료의 연구 활용 현황과 방법론적 과제

조용신 | 한국은행 경제연구원 경제안보연구실 부연구위원 | ycho32@bok.or.kr

I. 서론

북한경제 연구는 자료의 부족으로 일반 경제학의 정량분석 방법론을 적용하기 어렵다는 근본적인 어려움을 갖고 있다. 필자를 포함한 북한경제 연구자들은 자료가 없는데 북한경제에 대해 어떻게, 또 어떠한 연구를 할 수 있느냐는 질문을 자주 접하곤 한다. 잘 알려져 있다시피 북한은 세계에서 가장 폐쇄적인 국가로 인식되고 있으며, 많은 국가들이 발표하고 있는 경제총량(GDP, GNI 등), 물가, 실업률 등 대표적인 거시경제 통계는 물론 경제학의 틀 안에서 정량연구를 진행하기 위해 요구되는 신뢰도 있는 공식통계가 전무하다.

다행히도 국내 북한경제를 연구하고 있는 몇몇 기관에서 북한경제의 모니터링을 위한 기초자료를 추정 및 정리하여 발표하고 있다. 한국은행의 연도별 북한 GNI 추정치 및 분기별 물가 자료¹⁾, 한국무역협회의 월별 중국의 대북 수출입 자료²⁾, 한국토지주택연구원에서 노동신문 기사 중 건설 및 토지 관련 기사를 통계화하여 발간하고 있는 「북한건설개발 동향」과 「북한토지주택 리뷰」³⁾ 등이 있다. 또한 각 연구기관에서는 당 전원회의, 최고인민회의, 당대회 등 주요 정치적 이벤트 종료 후 노동신문의 결과 보고 기사를 분석하여 북한의 재정 상황 및 경제정책 현황을 추적하고 있다. 또한 최근에는 야간조도, 주간위성영상, 대기가스자료, 토지피복도 등 여러 대체 자료를 활용한 북한경제 분석도 시도되고 있다. 이러한 자료들은 최근 북한 거시경제의 흐름을 추적하는 데 상당히 유용하게 활용되고 있다.

1) 한국은행 경제통계시스템 (<https://ecos.bok.or.kr/#/>, 접속일: 2026.8.19.).
2) 한국무역협회(KITA) (<https://www.kita.net/>, 접속일: 2026.8.19.).
3) LH토지주택연구원 (<https://lhri.lh.or.kr/web/index.do>, 접속일: 2026.8.19.).

그러나 거시경제 측면으로만 북한경제를 이해하는 것에는 분명한 한계가 있다. 세계적으로 유례를 찾아보기 힘든 북한의 이원화 경제구조 때문이다. 정확한 비중은 알기 어려우나, 여러 관련 연구의 결과를 취합해 보면 시장부문은 적어도 일반 주민들의 일상 경제활동에 있어 공식부문 못지않은 비중을 차지하고 있는 것으로 파악된다(Kim and Song, 2008 등). 따라서 북한경제를 이해할 때 ‘비공식경제’로 통칭되는 시장부문에 대한 이해는 필수적이라고 볼 수 있다.

한편, 북한의 ‘비공식경제’는 그 용어에서 유추해 볼 수 있듯 일부 대형 종합시장을 제외하고는 아직 법적으로 인정받지 못하고 있으며 참여자들의 거래를 뒷받침하는 적절한 법적·제도적 장치, 사유권, 조세 및 재분배 제도, 공정경쟁 관련 제도 등도 마련되지 못하고 있다. 이러한 북한 시장부문의 특징은 북한경제 측면뿐 아니라 제도경제학 측면에서도 연구 가치가 높다고 할 수 있다. 그러나 신뢰할 만한 자료가 극도로 제한적인 북한에서, 시장이 갖는 비공식적 성격으로 인해 외부 연구자가 접근할 수 있는 정보의 양은 상당히 제한적인 상황이다. 국내외 북한전문 매체에서 현지 시장상황을 보도하고는 있으나, 북한의 시장경제를 보다 근본적이며 총체적으로 판단할 수 있는 자료는 찾기 힘든 실정이다.

이러한 상황에서 북한이탈주민 설문조사 자료는 시장경제의 동향, 작동기제, 파급효과 등을 파악하기에 가장 효과적인 자료라고 판단된다. 특히 설문조사 자료는 적절한 문항을 활용한다면 정량연구를 진행할 수 있어 통계적 검증이 가능하다는 점에서 북한경제 연구자들에게는 가치가 큰 자료이다.

북한이탈주민을 대상으로 하는 설문조사는 통일부, 서울대학교 통일평화연구원 등 정부 및 민간기관에서 실시된 바 있으며 대부분 북한의 비공식경제 부문에 대한 다양한 설문문항을 포함하고 있다. 물론 설문문항 중에는 비공식경제 외에도 북한 현지 주민들의 생활상과 인식을 다층적으로 살펴볼 수 있는 정치·사회·문화와 관련된 다양한 문항들도 포함되어 있다. 그러나 연구대상을 경제부문에 국한한다면, 설문조사 자료의 가장 큰 활용 가치는 비공식경제 부문 관련 문항에 있다고 해도 무리는 아닐 것이다. 본고는 현재까지 시행된 북한이탈주민 설문조사와 이를 활용하여 진행된 연구들을 소개하고, 설문조사를 활용하여 정량연구를 진행함에 있어 연구자가 반드시 고려해야 할 특징과 한계점을 정리해 보고자 한다.

II. 북한이탈주민 설문조사 및 활용 현황

북한이탈주민을 대상으로 하는 설문조사는 국내 여러 기관에서 시행되고 있다. 우선 현재 가장 많은 이탈주민을 대상으로 하는 설문조사는 통일부가 2010년부터 시행하고 있는 「북한 경제·사회 심층정보 수집」 사업이다. 이 설문조사를 활용하여 통일부는 매년 심층연구를 진행하고 있으나, 기본적으로 대외적으로는 공개하지 않고 있다. 다만, 2024년에 『북한 경제·사회 실태 인식보고서』라는 단행본이 출간되었는데, 동 보고서는 2013년부터 2022년까지의 북한이탈주민 대상 설문조사 결과를 정리하고 있다. 총 표본 수는 6,351명이며 경제부문에서는 경제 운영, 국영경제, 사경제 등 북한의 공식 및 비공식 경제 전반에 대한 설문조사의 문항별 결과를 담고 있다(이우영 외, 2024).

서울대학교 통일평화연구원의 경우 보기 드물게 정기적인 조사를 진행하여 시계열 데이터를 축적하고 있다. 2009년부터 시행된 설문조사는 「북한이탈주민 의식조사」와 「북한사회변동조사」라는 두 개의 설문으로 구성되어 있으며, 매년 직전 연도에 탈북한 이른바 ‘직행탈북자’만을 설문 대상으로 한다. 이러한 설문대상 표집 기준을 통해 설문대상자의 기억오류로 인한 응답의 부정확성 및 응답오류 우려를 최소화할 수 있어 북한 현지 상황의 변화를 잘 포착할 수 있다는 장점이 큰 설문조사이다. 아쉽게도 2020년 북중 국경 봉쇄 이후 북한 당국의 국경 경비 및 인프라가 대폭 향상되며 육로를 통한 탈북이 어렵게 되면서 탈북자 수가 급격히 줄어 현재는 설문조사가 중단되어 있는 상태이다. 하지만 2009년부터 2020년까지 총 1,241명을 대상으로 한 시계열 데이터가 축적되어 있는 만큼 연구 주제에 따라 연구 가치는 상당히 크다고 할 수 있다.

조사별 문항 구성을 살펴보면, 「북한이탈주민 의식조사」는 통일 인식, 대남 인식, 북한 실태 변화, 주변국에 대한 인식, 남한 적응 실태 등으로 구성되어 있으며, 「북한사회변동조사」는 북한의 실정 변화에 초점을 두고 북한에서의 식량, 의복, 시장활동, 사회복지, 정보통신 등에 대한 설문문항으로 구성되어 있다. 통일평화연구원은 원자료를 공개하고 있지는 않으나 설문조사가 실시된 연도마다 기초통계를 비롯하여 설문조사 결과에 대한 북한 전문가들의 심층연구 결과를 묶어 『북한주민 통일인식 2020』과 『북한사회변동』이라는 단행본을 출간하였다. 또한 2022년에는 김정은 정권 시기의 설문조사 결과들을 집대성하여 분석한 『김정은 집권 10년, 북한주민 통일인식』이라는 단행본을 출간하였다.

한편, 비정기적으로 실시된 설문조사 자료들도 존재한다. 먼저 서울대학교 통일평화연구원은 2020년에 기존 북한이탈주민 설문조사 프로그램이 종료된 이후 2024년에 약 500명의

북한이탈주민을 대상으로 단발성 설문조사를 실시하였다. 이 조사는 기존 설문조사의 대상자 선별 기준이었던 직전 연도 탈북이라는 조건을 적용하지 않고 국내 입국하여 거주하고 있는 모든 북한이탈주민을 대상으로 하였다. 이 조사는 정치·경제·사회문화·이주 및 남한 적응 실태 등 다양한 문항을 포함하고 있다는 점이 가장 큰 특징이다.

서울대학교 경제학부에서 실시한 북한이탈주민 대상 설문조사는 2004~05년, 2007년, 2009년, 2011년 등 총 네 차례에 걸쳐 시행됐다. 총 1,152명을 조사하였으며, 소그룹 단위의 갱 서베이(gang survey) 방식으로 진행되었다(Kim and Kim, 2016). 이 설문조사는 직업, 노동공급, 소득, 지출, 비공식경제 참여 여부, 비공식경제 참여 유형 등 공식 및 비공식 경제 부문 전반에 대한 문항들을 포함하고 있다. 특히 이 자료는 응답자의 탈북 의도 및 실제 탈북 과정에서 발생할 수 있는 교란을 통제하기 위하여 탈북 1년 전 시점의 경제생활을 기준으로 답하도록 설계하는 등 응답오류를 최소화하고자 하였다.

마지막으로 개별 연구자가 설문조사를 시행한 사례도 있다. 국방대학교의 김택빈 교수와 미국 University of North Texas의 존 이시아마(John Ishiyama) 교수가 구축한 설문조사는 2022년에 554명을 대상으로 진행되었다(Ishiyama and Kim, 2024). 특히 이 설문조사의 경우 세계가치관조사(World Value Survey)의 문항을 기초로 하되, 개별 문항을 북한이탈주민의 맥락에 맞게 조정하여 국제 비교가 가능하도록 설계하였다는 점에서 분석 대상 자료로서의 강점이 있다. 설문문항은 북한에서의 정치사회화 경험, 탈북 동기, 남한 재정착 및 재사회화 경험, 남한에서의 투표 행태, 투표 외 정치참여 행태 등 주로 정치·사회 주제 중심의 문항으로 구성되었다.

〈표 1〉 북한이탈주민 대상 설문조사 현황

설계자 및 기관	조사기간	표본 수	정기조사 여부
통일부 ¹⁾	2010~22년	6,351명	정기
서울대학교 통일평화연구원	2009~20년	1,241명	
	2024년	500명	
서울대학교 경제학부	2004~05년, 2007년, 2009년, 2011년	1,152명	비정기
John Ishiyama · 김택빈	2022년	554명	

주: 1) 통일부 조사는 비공개 조사로 자료가 공개되어 있지 않아 조사기간 및 표본 수는 통일부에서 공식 발간한 『북한 경제·사회 실태 인식보고서』의 2023년 조사를 기준으로 함.

지금까지 소개한 설문조사 자료를 활용하여 북한의 정치·경제·사회문화 등 전 분야에 걸쳐 많은 연구들이 이루어져 왔다. 그중 경제부문에 초점을 맞춰 설문조사 자료를 정량분석한 몇 개의 학술연구들을 소개하고자 한다. 설문조사를 활용한 북한경제 연구는 주로 비공식경제 부문에 집중되어 왔는데, 이는 비공식경제의 경우 설문조사 자료와 같은 1차 증언자료가 아니면 파악하기 힘들기 때문일 것이다.

Kim and Song(2008)은 서울대학교 경제학부 설문조사 자료를 활용하여 북한의 전반적인 비공식경제 규모 추정을 시도하였다. 그 결과, 비공식경제 소득이 전체 가구소득의 78%를 차지하는 것으로 나타났으며, 응답자의 70% 이상이 비공식 경제활동에 참여한 경험을 보고하였다. 특히 비공식 경제활동에 참여한 경험이 있다는 응답자 중 공식경제 종사자도 상당수를 차지하여 공식과 비공식 경제를 겸업하고 있는 주민의 비중도 상당할 것으로 추정하였다. 또한 도구변수(Instrumental Variable)를 활용한 분석 결과, 텃밭·가축사육형 활동은 공식부문 노동공급에 유의한 영향이 없었던 반면, 노점상·생산판매·밀수 등의 비공식 경제활동은 공식부문 노동시간을 유의하게 감소시켜 체제대체적(system-substituting) 활동인 것으로 나타나는 등 비공식부문 노동의 공식부문 노동 대체 가능성이 제기되었다. 나아가 Kim and Kim(2016)은 같은 자료를 활용하여 1996~2001년 기간 대비 2007~09년 응답자 중 서비스업(노점·운송·밀수)과 제조업(소비재 생산) 부문 비공식경제 참여자의 비중이 증가한 것으로 보고한 바 있다.

1990년대 중후반 극심한 경제난 이후 급속히 확산된 북한의 시장화에 대한 북한 당국의 정책적 대응은 일관성이 없었던 것으로 연구된다. 즉, 시장화 과정에서 친시장화 정책과 시장 억압 정책이 번갈아 시행되었으며, 현재까지도 대형 종합시장을 제외한 대부분의 시장을 법적으로 인정하지 않고 있다. 시장은 어디까지나 ‘비공식’의 형태로 당국의 묵인하에 운용되고 있다. 이러한 비공식적 성격 때문에 시장 참여자들은 대부분의 소득이 시장에서 발생함에도 불구하고 당국의 단속에 노출되어 있다. 당국의 단속을 피해 시장활동을 영위하기 위해서는 뇌물이 필수적인 것으로 알려져 있다. Kim and Koh(2011)는 서울대학교 경제학부 설문자료 중 2009년 3차 조사를 통하여 응답자의 절반가량(227명 중 112명)이 뇌물을 준 경험이 있으며 뇌물 지출이 전체 지출의 약 15.3%를 차지한다는 사실을 보고하였다. 또한 비공식경제 참여와 뇌물 제공 간에는 유의한 양(+)의 관계가 있는 반면, 뇌물 제공과 공식부문 노동시간 간에는 유의한 음(-)의 관계가 있다는 결과를 도출하였다. 이를 통하여 광범위한 비공식 경제활동은 단순히 경제적 생존 전략에 그치지 않고, 뇌물을 매개로 국가(사회주의 체제)의 공식적 통제·계획경제 시스템을 잠식하고 있다는 점을 유추해 볼 수 있다.

한편, 비공식 경제활동이 주민들의 정치·경제 관련 인식을 어떻게 변화시키는지 연구한 논문들도 있다. Kim and Cho(2025)는 서울대학교 통일평화연구원 설문자료를 분석하여 뇌물의 경험이 최고지도자에 대한 부정적 인식보다는 경제정책의 실무적 책임이 있는 당 지도부에 대해 부정적 인식을 형성한다고 주장한다. 경제 인식 측면에서 시장은 주민들의 경제적 불평등에 대한 인식에 영향을 미치는 것으로 보인다. Cho and Kim(2025)은 같은 설문조사 자료를 통하여 비공식 경제활동 참여 경험이 있는 응답자가 그렇지 않은 응답자 대비 사회 전체의 경제적 불평등도가 높다고 인식한다는 점을 지적한다. 이러한 결과는 시장에서의 경제활동이 참여자들에게 비참여자에 비하여 보다 많은 사회적 경험을 쌓을 수 있는 기회를 제공하며, 이를 통하여 참여자들이 경제적 불평등과 같은 사회 전체의 거시적인 상황을 평가할 수 있는 여러 정보를 제공받기 때문으로 해석된다. 또한 이렇게 경제적 불평등도가 높다고 인식하는 응답자일수록 최고지도자에 대한 부정적 인식을 형성할 가능성이 높다는 점도 보고한다.

북한이탈주민 설문조사 자료는 이렇듯 거시적 통계자료로 파악하기 어려운 북한 비공식경제의 규모, 작동 기제, 사회적 파급효과 등 여러 핵심 주제에 대한 정량분석을 가능케 하는 활용 가치가 높은 자료이다. 앞서 살펴본 연구들이 공통적으로 보여주듯, 이는 체제를 직접 경험한 당사자의 1인칭 증언을 자료화했기 때문에 가능한 성과이다.

그러나 바로 이 증언자료로서의 성격(응답자의 기억에 의존한 회고적 응답, 탈북이라는 특수한 경험을 공유한 집단에 국한된 표본, 상대적으로 작은 표본 규모 등)은 동시에 정량분석을 진행하는 연구자가 반드시 마주할 수밖에 없는 방법론적 한계이기도 하다. 다음 절에서는 이러한 한계의 구체적 내용과 이를 완화하기 위해 기존 연구들이 취해 온 대응 방식을 살펴본다.

Ⅲ. 북한이탈주민 설문조사 자료의 한계점과 대응방안

북한이탈주민 설문조사 자료는 그동안 의미 있는 연구 성과를 도출하는 데 기여했을 뿐만 아니라, 향후 수행 가능한 다양한 잠재적 연구주제를 제기한다는 점에서도 그 의의가 크다. 그러나 자료의 한계점도 분명하여 연구자의 특별한 주의 및 고민이 요구되는 자료이기도 하다. 앞서 소개한 북한이탈주민 설문조사를 활용한 연구들은 주제와 자료의 특성을 고려하여 OLS, Logit, Probit 등 회귀분석 방법론을 주된 분석방법론으로 채용하고 있으나, 그에 못지않게 설문조사 자료가 갖는 통계적 불안정성에 대응하기 위하여 다양한 강건성 검증 방법에

많은 지면을 할애하였다. 이번 장에서는 모든 이탈주민 설문조사 자료가 갖고 있는 가장 중요한 한계점들을 소개하고 이에 대응하기 위하여 연구자들이 사용한 방법들을 논의해 보고자 한다.

현재까지 시행되었거나 시행 중인 모든 북한이탈주민 설문조사는 표본의 표집방법이 무작위적이지 않다. 즉, 표본선택편향(sample selection bias) 문제를 안고 있다. 이 점은 설문조사 자료가 갖고 있는 한계점 중 가장 근본적이고 해결하기 쉽지 않은 문제이다. 북한이탈주민은 북한을 떠난 집단이라는 점에서 북한 전체의 모집단과 여러 면에서 다른 특징을 보일 수 있다. 이런 차이점은 이탈주민이 북한을 떠나고자 하였던 이유에 기인한다. 크게 경제적·정치적 이유 외에도 가족 재결합이나 자녀 세대에 대한 고려로 나눌 수 있는데, <표 2>의 서울대학교 통일평화연구원 설문조사 자료에 따르면 ‘경제적 어려움’은 16.4%, ‘북한 정권에 대한 불만·정치적 탄압(31.5%)’과 ‘자유체제에 대한 동경(18.8%)’ 등 정치적 이유는 50.3%를 차지한다. 두 경우 모두 북한 내 전체 모집단과 북한이탈주민이라는 표본의 특성을 구분 짓게 한다. 예를 들면 경제적인 어려움이 극심하여 탈북을 결심한 이탈주민 집단은 북한주민 전체의 경제 수준을 대표하지 못하는 소수의 인원이며, 정치적 탄압으로 탈북이라는 선택을 하게 된 집단은 북한 전체에서는 극히 예외적인 집단일 것이다. 이 근본적인 문제는 해당 집단을 분석한 회귀분석 결과의 의미를 북한의 일반 주민의 경향성으로 확장하기 어렵게 제한한다. 때문에 이탈주민 설문조사를 기반으로 하는 많은 연구에서 이 문제에 대해 독자들로 하여금 결과 해석에 주의가 필요하다는 점을 명시하고 있다.

<표 2> 탈북의 이유(2018~20년)

Q: 귀하께서 탈북을 결심하게 된 가장 큰 원인은 무엇입니까?		
	표본 수(명)	비중(%)
1. 북한 정권에 대한 불만·정치적 탄압	94	31.5
2. 경제적 어려움	49	16.4
3. 먼저 탈북한 가족의 권유	63	21.1
4. 자유체제에 대한 동경	56	18.8
5. 자식에게 더 나은 기회 부여	36	12.1
합계	298	100.0

자료: 서울대학교 통일평화연구원, 「북한주민 통일의식 2020」.

이 표본선택편향은 완전히 해결하기는 어려운 문제이다. 다만, 표본선택편향으로 인한 결과의 편향성을 완화하는 방법은 있을 수 있다. 예를 들어 앞서 언급한 탈북의 이유가 종속변수에 영향을 미치지 않도록 부표본(subsample)으로 나누어 분석을 진행하는 방법도 고려해 볼 수 있다. Cho and Kim(2025)은 주관적 경제적 불평등도가 최고지도자의 인식에 미치는 영향을 분석할 때 정치적 이유로 탈북한 집단, 즉 애초에 북한의 정권과 지도자에 대한 부정적 인식을 갖고 있었던 표본을 제외하여 분석을 진행하는 방식으로 표본선택편향에 대한 우려를 완화하고자 하였다. 이처럼 연구주제별 종속변수의 성격에 따라 표본 중 일정 집단을 배제하고 분석을 진행함으로써 독자들이 표본선택편향에 따른 편향 정도를 가늠해 볼 수 있도록 하는 방법도 고려해 볼 만하다. 그러나 앞서 언급하였듯 이러한 방법이 표본선택에 기인한 자료의 근본적 문제를 완전히 해결하지는 못한다는 점도 유념해야 할 것이다.

북한은 사회주의 사회로 개인의 다양성보다는 집단의 통일성을 강조하고 기초 교육 과정부터 사상 및 사유를 체계적으로 통제하는 체제이다. 이러한 체제적 특성으로 인해 일반적인 민주주의 시장경제 국가에 비해 정치체제나 최고지도자에 대한 인식과 같은 태도 변수(attitudinal variable)에 있어서는 주민들 간 인식의 분산 자체가 상대적으로 크지 않을 가능성이 있다. 이 경우 표본선택편향으로 인해 특정 성향의 응답자가 표본에 과대 혹은 과소 대표되더라도, 종속변수의 분산이 애초에 제한적이기 때문에 회귀분석 결과의 방향성이나 유의성이 크게 훼손되지 않을 수 있다. 다만, 이러한 논리는 인식·태도 관련 변수에 한정하여 조심스럽게 적용할 수 있는 것으로, 비공식경제 참여 여부나 소득, 노동시간과 같은 경제적 행태 변수(behavioral variable)에는 그대로 적용하기 어렵다. 개인의 경제활동은 거주지역, 성별, 가구 구성, 직업 등 관찰 가능한 조건에 따라 이질적으로 나타나는 것이 일반적이며, 앞서 소개한 연구들에서도 이러한 변수들에 따라 비공식경제 참여 양상이 유의하게 다르게 나타난다는 점을 확인한 바 있다. 또한 권위주의 체제하에서는 주민들이 실제 인식과 다르게 순응적인 태도를 표출하는 이른바 선호위장(preference falsification) 현상이 나타날 수 있다는 점도 유의할 필요가 있다(Kuran, 1995). 이 경우 인식의 동질성은 실제 동질성이라기보다 체제의 억압에 따른 응답 왜곡의 결과일 수 있어, 오히려 자료의 타당성(validity) 문제로 이어질 가능성도 배제하기 어렵다. 따라서 연구자는 종속변수가 태도·인식 관련 변수인지 혹은 경제적 행태 변수인지를 먼저 구분한 뒤, 표본선택편향이 분석 결과에 미치는 영향의 정도를 신중하게 가늠할 필요가 있다.

표본선택편향의 연장선상에서 다음으로 살펴볼 문제는 성별과 지역 비중의 편중이다. 통일부의 북한이탈주민 공식통계에 따르면 2026년 6월 말 기준 총 34,536명의 북한이탈주민이

남한에 정착하고 있는데, 이 가운데 여성 비중이 72.2%이며, 접경지역인 양강도와 함경북도 출신 비중은 76.2%로 성비와 지역 비중 모두 크게 편중되어 있다. 이에 따라 전체 북한이탈주민을 모집단으로 하는 설문조사도 조사기관별 약간의 차이는 있으나 기본적으로 같은 문제를 안고 있다. 지역 편중의 경우 그동안 탈북의 주요 루트가 육로를 통한 탈북이었던 점에서 북중 접경지역인 양강도와 함경북도 출신 이탈주민의 비중이 높은 것으로 해석된다. 여성 비중의 경우 잘 알려져 있다시피 북한의 남성은 대부분 공식 직장이 있으며, 이로 인하여 당국의 감시가 보다 촘촘하기 때문에 이동의 자유가 여성 대비 훨씬 제한적이다. 반면, 여성은 대부분 시장에 참여하는 경우가 많아 정보 접근성이 좋으며, 당국의 감시가 비교적 느슨한 관계로 탈북의 난이도가 남성보다 낮을 가능성이 크다. 북한이탈주민 모집단의 편중된 성비는 경제활동의 유형과 당국의 감시 강도 차이에 기인한 것으로 추정된다.

〈표 3〉 성별·지역별 비중

(단위: %)

	여성 비중(%)	접경지역 비중(%)
통일부, ‘북한 경제·사회 실태 인식보고서’ ¹⁾	81.8	82.1
서울대학교 통일평화연구원, ‘북한주민 의식조사’ ²⁾	60.0	85.9
서울대학교, ‘경제학부 북한이탈주민 설문조사’ ³⁾	62.7	75.8

자료: 1) 「북한 경제·사회 실태 인식보고서」.
2) 「북한주민 통일의식 2020」.
3) Kim and Kim(2016).

이에 더해 시계열적 분석 측면에서의 어려움도 있다. 대부분의 이탈주민 설문조사는 조사 가능한 표본 수의 한계로 인하여 일반적인 설문조사에서 표집방법으로 채택하는 다단계층화표집을 시행하기 어렵다. 즉, 설문조사 시점별 인구통계학적 구성에 차이가 발생할 수밖에 없다. 예컨대 서울대학교 통일평화연구원 설문조사 자료의 인구통계의 시계열적 흐름을 살펴보면(표 4), 조사연도마다 성별, 교육 수준, 당원 여부 등의 비중이 다를 수 있다.

〈표 4〉 연도별 주요 인구통계 비중 추이

(단위: %)

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
여성 비중	65.7	55.9	56.1	66.4	61.0	61.6	47.0	56.3	64.7	66.1
고등 중졸	66.7	67.5	71.8	62.6	61.4	74.6	72.5	73.6	70.7	70.9
당원 여부	9.9	14.4	14.4	13.5	17.9	18.8	18.3	14.0	11.3	13.1

자료: 「북한주민 통일의식 2020」.

이러한 인구통계적 편중과 시계열적 비일관성 문제에 대응하기 위하여 관련 연구들은 크게 두 가지 방법론을 활용해 왔다. 리샘플링(resampling)과 성향점수매칭(propensity score matching, PSM)이 그것이다.

먼저 리샘플링은 비교 대상이 되는 두 집단, 예컨대 서로 다른 조사연도나 서로 다른 시기의 응답자 집단 간 인구통계학적 구성을 최대한 유사하게 맞추기 위하여 각 집단에서 표본을 재추출하는 방법이다. 만약 두 시기 표본 간 성별, 거주지역, 주관적 계층인식, 직업 등 주요 인구통계 변수의 구성비가 서로 다르다면, 이러한 차이가 완화되도록 표본을 반복적으로 재추출하여 두 집단의 구성비를 근사시킨 뒤 분석을 진행하는 방식이다. 아래 〈표 5〉는 Cho(2023)의 리샘플링 처리 결과를 보여준다. 동 연구는 2014~15년 북한경제가 경험한 외생적 경제충격⁴⁾ 이전(1기)과 이후(2기) 북한이탈주민의 비공식 소득 양극화 변화를 측정하였는데, 이를 위해서 두 기간 간 여성 비중, 함경북도 출신 비중, 주관적 사회계층 문항 중층 응답 비중, 시장에서 소매 장사를 해본 경험 비중 등의 표본 특성 차이를 최소화하려는 목적으로 리샘플링 처리를 시도하였다. 그 결과 〈표 5〉에서 확인할 수 있듯 두 기간 간 표본 특성 차이가 다소 줄어들고 동시에 표본 수는 감소한 것을 확인할 수 있다.

4) 2014~15년 기간 북한은 심각한 가뭄으로 인한 식량 생산량 감소와 전력 생산 부진을 경험했으며, 2015년 중국 경기가 둔화되고 국제 원자재 가격이 하락하며 주력 수출액이 감소한 바 있다.

〈표 5〉 리샘플링 처리 결과 예시

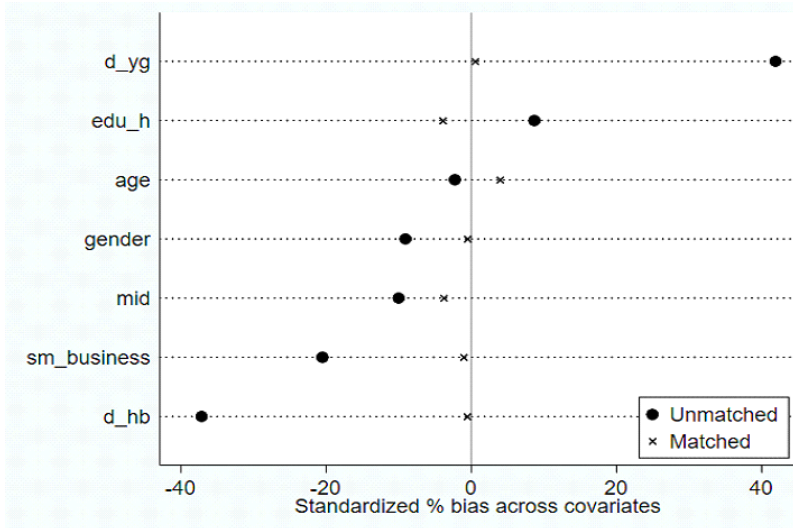
	처리 전(%)		처리 후(%)	
	1기	2기	1기	2기
여성 비중	1기	60.4	1기	61.5
	2기	58.3	2기	60.0
함경북도 출신 비중	1기	41.5	1기	39.0
	2기	25.3	2기	33.0
주관적 사회계층(중층) 비중	1기	73.8	1기	74.0
	2기	73.9	2기	74.0
시장의 소매 장사 경험 비중	1기	45.9	1기	44.0
	2기	35.7	2기	40.5
표본 수	1기	366	1기	200
	2기	415	2기	200

자료: Cho(2023).

이 방법은 절차가 비교적 단순하고 직관적이어서 결과 해석이 용이하다는 장점이 있다. 다만, 리샘플링은 어디까지나 표본 내에서의 구성비를 사후적으로 조정하는 것에 불과하므로, 애초에 표본 자체가 모집단을 대표하지 못한다는 근본적인 문제, 즉 표본선택편향 자체를 해소하지는 못한다는 한계를 지닌다.

두 번째로 PSM은 관찰 가능한 특성이 유사한 응답자끼리 짝을 지어 비교집단을 구성함으로써, 두 집단의 구성을 통계적으로 보다 정교하게 맞추는 방법이다. 성별, 연령, 거주지역, 교육 수준, 직업 등의 특성을 바탕으로 각 응답자가 특정 집단, 예컨대 특정 시기나 특정조사에 속할 확률(성향점수)을 추정한 뒤, 이 확률이 비슷한 응답자끼리 매칭하여 비교표본을 구성하는 방식이다. Kim and Kim(2016)은 서울대학교 경제학부 설문조사 자료를 활용한 분석에서, 2005~09년 탈북자와 1996~2004년 탈북자 간 인구통계 구성의 차이를 통제하기 위해 PSM을 적용하였으며, 매칭 이후 두 시기 표본 간 거주지역·학력 등 주요 변수의 편향이 크게 감소하였음을 확인한 바 있다. Cho(2023) 또한 PSM을 사용한 바 있다. 앞선 리샘플링의 경우와 같은 이유로 PSM을 시행하였으며 결과는 아래 [그림 1]과 같다. 그림은 PSM 전(unmatched)과 후(matched) 공변량별 표준화 편향 정도를 보여준다. 가로축은 두 집단 간 각 변수의 표준화 백분율 편향을 나타내며, 0에 가까울수록 두 집단 간 해당 변수의 구성이 유사함을 의미한다. 그림에서 확인할 수 있듯 PSM 시행 이후 양강도·함경북도 출신 비중, 고등중학교 졸업 비중, 평균 연령, 여성 비중, 시장에서의 소매 장사 경험 비중 등의 편향이 대폭 줄어든 것을 확인할 수 있다.

[그림 1] PSM 시행 결과 예시(시행 전후 표준화 편향)



주: 세로축의 d_yg과 d_hb는 양강도와 함경북도 출신 비중을 의미하며, edu_h는 고등학교 졸업 비중, age는 연령, gender는 성별, mid는 주관적 사회계층 문항 중응 응답 비중, sm_business는 시장 소매 장사 경험 비중을 각각 의미함.
자료: Cho(2023).

PSM은 리샘플링보다 통계적으로 엄밀한 절차를 거친다는 장점이 있으나, 매칭 조건에 부합하지 않는 표본이 다수 제외되면서 최종 분석표본의 규모가 크게 줄어들 수 있다는 한계가 있다. 특히 표본 수 자체가 제한적인 이탈주민 설문조사의 특성상, 매칭으로 인한 표본 축소는 분석 결과의 통계적 검정력(statistical power)을 약화시킬 위험이 있다.

요컨대 리샘플링과 PSM은 모두 비교 대상 집단 간 인구통계적 구성을 유사하게 맞추으로써 분석 결과가 인구통계적 차이에 의해 왜곡되었을 가능성을 완화하는 데에는 유용하지만, 두 방법 모두 표본 자체의 대표성 문제, 즉 근본적인 표본선택편향을 해소하는 방법은 아니라는 점에서 한계가 뚜렷하다. 따라서 이들 방법론을 활용한 연구 결과를 해석할 때에도, 어디까지나 표본 내 비교의 신뢰도를 높이는 보조적 수단이라는 점을 염두에 두어야 할 것이다.

마지막으로 응답자의 응답오류(response error) 가능성도 이탈주민 설문조사를 활용한 정량연구에서 반드시 고려해야 할 문제이다. 특히 비공식 소득과 같이 응답자가 정확한 액수를 기억하기 어렵거나, 응답 자체를 꺼릴 만한 민감한 항목의 경우 이러한 우려가 더욱 커진다. 비공식 경제활동은 그 성격상 당국에 노출되어서는 안 되는 활동이었기 때문에, 응답자가 실제 소득 규모를 축소하거나 반대로 부풀려 응답할 가능성을 배제하기 어렵다.

또한 대부분의 설문이 탈북 이후 상당한 시간이 지난 시점에 과거의 기억에 의존하여 진행되는 만큼, 단순한 기억의 오차로 인해 응답값이 실제와 다르게 보고될 가능성도 존재한다.

이러한 응답오류에 대응하기 위해서는 통계적 기법 이전에 응답 내용 자체에 대한 세심한 데이터 클리닝(data cleaning) 과정이 선행되어야 한다. 예컨대 비공식 소득 관련 문항의 경우, 해당 응답이 다른 관련 문항들과 논리적으로 부합하는지를 교차 확인하는 작업이 필요하다. 공식 직장이 없다고 응답한 사람이 공식 소득을 함께 보고하였거나, 반대로 비공식 경제활동에 참여하지 않았다고 응답하였음에도 불구하고 비공식 소득을 보고한 경우, 혹은 주된 소득원이 비공식부문이라고 답했음에도 공식 소득이 비공식 소득보다 크게 보고된 경우 등은 응답 간 정합성이 결여된 사례로 볼 수 있어 분석에서 제외하는 방식이 일반적으로 사용된다. 이 밖에도 스스로를 하위계층이라 인식한다고 응답하였음에도 불구하고 총소득이 지나치게 높게 보고된 경우와 같이, 주관적 인식과 객관적 응답값 간의 괴리가 뚜렷한 사례 역시 같은 방식으로 걸러낼 수 있다.

이와 더불어 극단값(outlier)에 대한 처리도 필요하다. 소수의 응답자가 보고한 비현실적으로 큰 소득값이 전체 분석 결과를 왜곡할 수 있으므로, 연도별 분포에서 일정 백분위수를 크게 상회하는 값은 통계적 이상치로 간주하여 제외하는 방식이 흔히 활용된다. 반대로 지나치게 낮은 값 역시 실질적인 경제활동 참여자로 보기 어려운 경우가 많아 필요에 따라 별도의 하한선을 두어 제외하는 것이 적절하다.

이러한 절차는 다소 수작업에 가까운 반복적인 검토를 요구하며, 어떤 기준을 적용하느냐에 따라 최종 분석표본의 구성이 달라질 수 있다는 점에서 연구자의 자의적 판단이 개입될 여지도 존재한다. 그럼에도 불구하고 응답오류 가능성이 상존하는 이탈주민 설문조사의 특성상, 이러한 문항 간 정합성 검토와 이상치 처리 과정은 분석 결과의 신뢰성을 확보하기 위한 필수적인 절차로 판단된다. 연구자는 가능한 한 임의 판단을 지양하고 명확하고 보수적인 기준을 세워 그 기준을 엄격하게 지키며 모든 과정을 독자가 확인할 수 있도록 투명하게 명시하여야 한다.

지금까지 살펴본 세 가지 한계점은 그 성격과 대응 가능성에 있어 분명한 차이를 보인다. 표본선택편향은 이탈주민이라는 표본 집단이 갖는 태생적 특성에서 비롯되는 문제로, 어떠한 통계적 기법으로도 완전히 해소하기 어려운 근본적인 한계이다. 반면, 성별·지역 편중과 시계열적 비일관성은 리샘플링이나 PSM과 같은 방법론을 통해 그 영향을 상당 부분 완화할 수 있으며, 응답오류 역시 세심한 데이터 클리닝 과정을 거치면 분석 결과의 신뢰성을 크게 높일 수 있다. 즉, 이탈주민 설문조사 자료가 안고 있는 한계는 그 정도에 따라 근본적으로

감내해야 하는 문제와, 연구자의 노력을 통해 상당 부분 극복할 수 있는 문제로 구분해서 접근할 필요가 있다. 결국 중요한 것은 이러한 한계의 존재를 정확히 이해하고 이에 맞는 적절한 방법론을 선택하며, 그 과정과 한계를 독자에게 투명하게 밝히는 연구자의 태도일 것이다.

IV. 결론

본고에서는 북한이탈주민을 대상으로 한 여러 설문조사 자료의 현황과 이를 활용한 대표적 연구 사례들을 살펴보고, 이러한 자료가 갖는 방법론적 한계와 이에 대한 대응방안을 논의하였다. 북한경제는 신뢰할 수 있는 공식통계가 사실상 부재하다는 근본적 제약을 안고 있으며, 특히 전체 경제활동에서 상당한 비중을 차지하는 비공식경제 부문은 거시통계만으로는 파악이 원천적으로 불가능하다. 이러한 상황에서 체제를 직접 경험한 당사자의 증언을 자료화한 북한이탈주민 설문조사는 정량분석이 가능한 사실상 유일한 미시 자료로서, 통일부, 서울대학교 통일평화연구원, 서울대학교 경제학부, 그리고 개별 연구자에 이르기까지 다양한 주체에 의해 축적되어 왔다. 이를 활용한 연구들은 비공식경제의 규모와 구조, 그 결과물로 북한 사회에 만연하게 된 뇌물 문제, 경제적 불평등 인식, 정치적 태도 변화 등 거시통계로는 접근할 수 없었던 북한 사회의 실질적 작동방식을 실증적으로 규명해 왔다.

다만, 이러한 자료는 표본선택편향, 성별·지역별 인구통계적 편중, 조사 시점별 구성의 비일관성, 응답오류 가능성 등 여러 방법론적 한계를 동시에 안고 있다. 이 중 표본선택편향은 탈북이라는 행위 자체가 갖는 본질적 특성에서 비롯되는 만큼 완전한 해소가 어려운 근본적 한계인 반면, 인구통계적 편중과 시계열적 비일관성은 리샘플링과 PSM을 통해, 응답오류는 세심한 데이터 클리닝을 통해 그 영향을 상당 부분 완화할 수 있다. 결국 이러한 한계의 성격을 명확히 구분하여 인식하고, 이에 상응하는 적절한 방법론을 선택하며, 그 과정을 투명하게 제시하는 것이 북한이탈주민 설문조사 자료를 연구하는 연구자에게 가장 중요한 부분이라고 할 수 있다.

한편, 이러한 논의는 앞으로 더욱 중요해질 것으로 보인다. 통일부 공식통계에 따르면 북한이탈주민의 연간 국내 입국 인원은 2020년 코로나19로 인한 복중 국경봉쇄 이후 229명(2020년), 63명(2021년), 67명(2022년)까지 급감하였다. 2023년과 2024년 들어 각각 196명, 236명으로 다소 회복되는 추세를 보이고는 있으나, 과거에 비하면 여전히 매우 낮은 수준이다.

이에 따라 서울대학교 통일평화연구원이 매년 진행해 오던 「북한이탈주민 의식조사」는 조사대상이었던 타국 비경유 직행탈북자 확보가 어려워지며 2020년을 끝으로 중단되었으며, 이후 2024년에 실시된 조사 역시 기존의 표집 기준을 적용하지 못하고 국내 정착 이탈주민 전체를 대상으로 한 단발성 조사로 전환할 수밖에 없었다. 이는 비단 통일평화연구원만의 문제가 아니라, 신규 입국 탈북민을 표집 대상으로 삼는 모든 조사주체가 공통적으로 마주하고 있는 현실이다. 즉, 북한 현지 상황의 변화를 새롭게 반영한 시계열 자료의 축적은 당분간 기대하기 어려운 상황이다.

그럼에도 불구하고 지금까지 축적된 북한이탈주민 설문조사 자료는 여전히 대체하기 어려운 가치를 지닌다. 첫째, 이 자료들은 1990년대 중반 고난의 행군 이후 북한 사회가 계획경제에서 시장경제로 급격히 이행해 온 과정을 담고 있는 사실상 유일한 정량적 기록이다. 둘째, 앞서 살펴본 연구들에서 확인하였듯 이 자료는 단순한 현황 파악을 넘어 정책 변화에 따른 주민들의 행태 변화, 체제에 대한 인식 형성 기제 등 북한 경제·사회의 작동 원리, 특히 바공식경제의 면면을 살펴보는 데에도 기여해 왔으며, 이러한 분석적 자산은 향후 통일 이후 북한경제를 이해하고 정책을 설계하는 데에도 유용한 기초자료로 활용될 수 있다. 셋째, 신규 표본 확보가 제한적인 현재의 상황은 오히려 기존에 축적된 자료를 보다 정교하게 재분석하고 서로 다른 기관의 자료를 상호 연계하여 활용하는 방식의 중요성을 부각시킨다. 이러한 점에서 통일부와 서울대학교 통일평화연구원 등 조사기관이 원자료 공개 범위를 확대하는 것도 향후 연구의 지속가능성을 위해 함께 고민해야 할 과제일 것이다.

북한이탈주민 설문조사 자료는 신규 정보 유입이 제한된 현재의 상황 속에서도, 앞서 논의한 방법론적 한계를 충분히 인지하고 신중하게 접근한다면 북한 경제와 사회를 이해하는 데 있어 여전히 가장 효과적이고 대체하기 어려운 자료가 될 수 있다. 최근에는 야간조도(nighttime light)나 위성영상과 같은 지리정보(GIS) 자료를 활용하여 공식통계의 공백을 메우려는 시도들도 활발히 이루어지고 있다. 이러한 대체자료는 거시경제 부문을 객관적 데이터를 기반으로 분석할 수 있다는 점에서는 강점을 갖는다. 하지만 그 이면에서 작동하는 주민들의 미시적 경제행위와 동기, 인식까지 담아내지는 못한다는 점에서 북한이탈주민 설문조사 자료의 연구 가치가 결코 뒤지지 않는 만큼, 학계의 지속적인 관심과 투자가 요구된다.

참고문헌

- 김학재 · 김병로 · 문인철 · 엄현숙 · 조동준 · 천경효 · 임경훈 · 최현정, 『북한주민 통일의식 2020』, 통일학연구 51, 서울대학교 통일평화연구원, 2021.
- 이우영 · 고광영 · 김성경 · 박주화 · 양문수 · 유영수 · 이철수 · 조용신, 『북한 경제 · 사회 실태 인식보고서』, 통일부, 2024.
- Cho, Yongshin, “Studies on Economic Classes of North Korea and Unification Perception of the South Koreans,” PhD diss., Seoul National University, 2023.
- Cho, Yongshin and Taekbin Kim, “Seeing Disparity: Marketization and Perceived Inequality in North Korea,” *World Development*, Vol. 196, 107179, 2025.
- Ishiyama, John and Taekbin Kim, “Attitudinal Legacies of Dictatorship: How Premigratory Experiences Affect North Korean Defectors' Attitudes on Authority and Nationalism,” *Asian Survey*, Vol. 64, No. 4, 2024, pp.605~634.
- Kim, Byung-Yeon and Min Jung Kim, “The Evolution of the Informal Economy in North Korea,” *Seoul Journal of Economics*, Vol. 29, No. 4, 2016, pp.457~480.
- Kim, Byung-Yeon and Yu Mi Koh, “The Informal Economy and Bribery in North Korea,” *Asian Economic Papers*, Vol. 10, No. 3, 2011, pp.104~117.
- Kim, Byung-Yeon and Dongho Song, “The Participation of North Korean Households in the Informal Economy: Size, Determinants, and Effect,” *Seoul Journal of Economics*, Vol. 21, No. 2, 2008, pp.361~385.
- Kim, Taekbin and Yongshin Cho, “Effects of Everyday Corruption on Perceptions of Authoritarian Leadership: Evidence from North Korea,” *Problems of Post-Communism*, Vol. 72, No. 5, 2025, pp.464~475.
- Kuran, Timur, *Private Truths, Public Lies: The Social Consequences of Preference Falsification*, Cambridge, MA: Harvard University Press, 1995.
- 〈자료〉
- 통일부 주요사업통계(<https://www.unikorea.go.kr/web/unikorea/contents/statistics>, 접속일: 2026. 8. 19).
- 한국은행 경제통계시스템 (<https://ecos.bok.or.kr/#/>, 접속일: 2026.8.19.).

한국무역협회(KITA) (<https://www.kita.net/>, 접속일: 2026.8.19.).

LH토지주택연구원 (<https://lhri.lh.or.kr/web/index.do>, 접속일: 2026.8.19.).