

사업체패널조사의 표본이탈 및 무응답 편향 현황*

김기민**

본 연구는 2015년에 새롭게 추출한 사업체패널조사 패널표본을 활용하여 표본이탈 현황을 분석하고 표본이탈이 대표성과 추정치 편향에 미치는 영향을 분석함으로써 향후 사업체패널조사의 지속가능성을 검토하였다.

2015년 표본의 응답률은 2차 조사 80.5%, 3차 조사 78.5%, 4차 조사 72.0%로 2005년 표본의 응답률보다 전반적으로 높았고, 국내·외 다른 조사에 비해서도 높게 나타났다. 2015년 표본의 무응답 편향은 2차 조사 2.26%, 3차 조사 3.09%, 4차 조사 2.68%로 크지 않았다. 또한 무응답 편향을 검정할 수 있는 하위집단 응답률의 변동계수는 2~4차 조사 동안 10% 미만을 유지했고 무응답 가중치의 분산 또한 크게 증가하지 않았다. 이를 통해 2015년 표본은 4차례의 조사가 진행되는 동안 패널표본의 대표성이 크게 훼손되지 않았음을 확인하였다. Monte Carlo simulation을 활용하여 향후 실시될 2015년 표본의 조사 응답 상황을 예측한 결과, 사업체패널조사는 향후 2차례 정도 조사를 하여도 데이터의 품질(quality)을 잘 관리할 수 있을 것으로 나타났다.

본 연구를 통해 사업체패널조사는 2015년 신규표본을 구축한 이후 4차 조사까지 잘 유지되고 있음을 확인하였다. 4차례의 조사수행 상황과 모의실험 결과를 고려하면 앞으로 2차례 정도는 패널 조사가 가능할 것으로 보이나, 향후 특정 집단에서 무응답률이 증가하면 현재의 예측과 다른 결과가 나타날 수 있다. 이를 예방하기 위해서는 대규모 사업체 등과 같이 응답 확률이 낮은 특성을 보이는 사업체의 참여를 높일 수 있는 방안을 모색할 필요가 있다.

주제어: 사업체패널조사, 표본이탈, 응답률, 무응답 편향, 몬테카를로 모의실험

* 본 논문은 한국노동연구원 ‘패널자료 품질개선 연구(XIII) 연구보고서’로 발표된 것을 수정·보완한 것입니다. 본 논문을 위해 유익한 조언을 해주신 익명의 두 분 심사자에게 깊은 감사의 뜻을 전합니다.

** 한국노동연구원 전문위원(kimin1104@kli.re.kr).

I. 서론

패널조사는 조사 목적에 맞게 추출된 패널표본을 대상으로 여러 시점에 걸쳐 추적하는 조사이다. 패널조사는 동일 표본을 반복적으로 조사하기 때문에 횡단면 조사와 시계열 조사의 장점을 동시에 갖는다. 그러나 패널조사는 조사가 반복됨에 따라 표본이탈(sample attrition)¹⁾이 누적적으로 발생한다는 문제점을 갖고 있다. 이러한 표본이탈은 추정치의 신뢰성을 떨어뜨리며 심해지는 경우 패널조사의 지속가능성이 위협받는다. 더욱이 표본이탈이 무작위적이지 않고 체계적(non-ignorable)인 경우 표본 대표성이 훼손된다. 본 연구는 한국노동연구원에서 수행하는 사업체패널 조사를 활용하여 사업체 단위 패널조사의 패널이탈 현황 등, 자료의 품질을 분석하고 향후 조사의 지속가능성을 알아보고자 한다.

한국노동연구원의 사업체패널조사는 2005년에 1,749개 표본 사업체를 구축하여 2013년까지 5차례의 패널조사를 실시하였다. 최초 조사 이후 4차례에 걸친 추적조사의 과정에서 거절, 휴·폐업 등으로 패널표본 사업체의 탈락(attrition)이 지속적으로 진행되면서 표본 대표성을 확보하기 위해 매 차수마다 이탈표본에 대해 표본설계 기준에 따른 표본대체를 실시하였다. 누적된 표본대체와 사업체 특성변화로 인해 사업체패널조사의 지속가능성은 물론 동태적 변화를 파악함에 있어서의 적절성 문제가 제기되었다. 이를 개선하고자 2015년 조사부터 신규표본을 추가하여 3,400여개 사업체를 대상으로 조사를 완료하였다. 과거 조사에 비해 약 2배 정도 많은 표본 수를 확충하였을 뿐 아니라 표본대체로 인해 모집단 정의가 불명확해지는 것 또한 방지하고자 2015년 조사부터 표본대체를 실시하지 않았다.

신규표본을 대상으로 2015년에 시작된 사업체패널조사는 현재 4차례의 패널조사를 완료하였다. 2015년의 신규표본 또한 4차례의 조사가 진행되면서 과거와 마찬가지로 패널이탈이 발생하였다. 본 연구는 4차례에 걸친 조사 동안의 표본이탈 현황과 무응답 편향 분석을 통해 사업체패널조사의 표본 상태를 점검하고 모의실험을 활용하여 향후 몇 차례의 조사가 더 가능할 수 있는지 기능해 보고자 한다.

1) 패널표본을 구축하는 첫 조사(1차 조사)의 단위 무응답은 표본이탈에서 제외되며, 2차 조사부터 표본에 대한 추적 실패나 응답 거절로 인해 발생한 단위 무응답을 표본이탈로 정의한다.

본 연구는 다음과 같이 구성된다. 우선 제2장은 패널자료의 이탈현황을 분석한 선행연구를 살펴본다. 제3장에서 자료를 소개하고, 제4장은 2015년에 설계된 신규 표본의 응답 현황 및 무응답 편향, 그리고 응답 성향을 분석한다. 제5장에서는 모의 실험을 통해 사업체패널조사의 지속가능성을 예측한다. 마지막으로 제6장에서 분석 결과를 요약하고 향후 사업체패널조사의 운영방안에 대해 논의한다.

II. 선행연구 검토

패널자료의 품질 유지를 위해 조사대상의 이탈을 최소화하는 것은 매우 중요하다. 표본이탈은 1998년 「*Journal of Human Resources*」에서 표본이탈에 관한 특집을 다루었을 정도로 패널자료에서 중요한 이슈이다(이상호 2005, p.69). 그러나 표본이탈에 관한 많은 연구들은 개인이나 가구 조사에 집중되어 있다.

사업체 혹은 기업을 대상으로 한 조사는 개인이나 가구 조사와는 다른 특성이 존재한다. 개인이나 가구의 무응답은 조사 대상 스스로의 참여 결정을 통해 이루어지지만, 사업체 조사의 무응답은 사업체 내 응답자가 참여를 거부한 것일 수도 있고 사업체의 정책에 따른 조사 거부일 수도 있다. 따라서 일반적인 가구나 개인에 대한 패널조사의 표본이탈과는 다른 특성이 나타날 수 있다. 즉 사업체 혹은 기업 조사의 표본이탈에 대한 연구에서 조사 응답자는 개인이지만 사업체는 개인이 아니기 때문에 이러한 차이점을 염두에 두고 분석해야 한다(Tomaskovic-Devey, Leiter, & Thompson, 1994). 또한 Petroni et al. (2004)는 사업체 조사의 무응답 편향은 모집단을 왜곡시킬 수 있어 가구 조사보다 더 큰 문제가 될 수 있음을 강조했다.

Hidiroglou, Drew & Gray(1993)는 기업 조사와 사회 조사(social survey)의 무응답을 줄이기 위한 조사 절차를 기술하였다. 기업 조사와 가구 조사를 명시적으로 비교한 것은 아니지만 기업 고유의 특성이 조사 무응답에 영향을 미친다고 보고했다. Tomaskovic-Devey, Leiter, & Thompson(1994)는 기업의 특성이 조사 참여에 영향을 미친다는 사실을 보고하고 이를 설명할 수 있는 이론적 모형을 제시하였다. 기업 조사에서 응답자의 참여는 응답자의 권한, 역량, 응답 동기에 의해 결정되며 또한 기업의 환경도 참여에 영향을 미친다고 보고했다. 따라서 기업 조사의 응답률을 높이기 위해서는 응답자 선정에 신중해야 하며 기업의 경영환경(규모, 경영체제 등)

을 파악하는 것이 도움이 될 것이라 제안하였다. 사업체 혹은 기업 조사의 표본이 탈 분석시 해당 조사의 응답 여부는 응답자가 속해 있는 조직의 상황에 의해 결정될 수 있기 때문에 개인이나 가구 조사와는 달리 응답자의 특성뿐 아니라 응답자가 속해 있는 조직의 특성 또한 반영해야 함을 제시하였다. 그러나 외국에서도 사업체나 기업 단위의 패널조사는 흔치 않기 때문에 기업이나 사업체의 무응답 연구는 주로 횡단면 조사 혹은 2차 정도의 짧은 패널자료를 활용한 분석에 그쳤다. 1990년대부터 시작된 독일의 IAB Establishment Panel Survey 자료가 어느 정도 축적된 후, 긴 기간에 대한 사업체 단위의 패널연구가 가능해졌다.

Janik & Kohaut(2012)는 독일의 IAB Establishment Panel Survey 자료를 분석하여 사업체의 응답자가 관련 사업체에 대한 정보를 제공할 권한이 있고, 설문에 신뢰할 수 있는 응답을 제공할 수 있는 역량이 있으며, 경영 측면에서 사업체가 해당 조사에 관심이 있는 경우 조사를 거절할 확률이 적음을 보였다. 또한 응답자와 조사원 중 한 사람이라도 변경되는 경우²⁾ 참여 확률이 감소함을 보여, 응답자와 조사원 간의 상호작용이 중요함을 보였다. Earp et al.(2018)은 회귀 나무(regression trees) 분석방법을 활용하여 사업체 종단조사의 무응답을 분석하였는데, 고용 규모, 단독/다수 사업장, 소유경영체제, 업종 등이 무응답과 관련됨을 확인하였다. König et al.(2021)과 König & Sakshaug(2023)은 독일의 IAB Establishment Panel Survey 자료를 활용하여 17년 동안(2001년부터 2017년)의 사업체 조사에서의 무응답률 및 무응답 편향의 추세, 무응답률과 무응답 편향과의 관계, 그리고 사업체와 노동자 특성이 조사 참여 여부에 미치는 영향에 대해 분석하였다. 사업체 조사의 응답률은 점차적으로 감소하는 추세를 보였는데, 거부 편향은 증가하는 반면 비접촉 편향은 동기간 동안 고르게 변동함을 확인하였다. 또한 해당 기간 동안 무응답 편향이 증가함을 보여, 무응답률과 무응답 편향은 작지만 양의 상관관계임을 확인하였다. 추가적으로 1,000명 이상의 대규모 사업체는 소규모 사업체에 비해 조사의 참여 가능성이 낮고, 인구가 적은 지역에 위치한 사업체는 인구가 많은 지역에 위치한 사업체에 비해 조사 참여 가능성이 높았다. 이 연구는 응답률과 더불어 거절율과 비접촉율을 고려하고 무응답 편향을 노동력 구성과 사업체 특성별로 구분하여 제시하였는데, 본 분석은 이 연구의 분석방법을 차용하였다.

2) 사업체 단위의 패널조사는 개인이나 가구 단위의 패널조사와 달리 응답자가 변경될 수 있다. 인사이동 등을 통해 해당 직무의 응답자가 변경되는 경우 이에 해당된다. 응답자 변경으로 인한 자료의 품질에 미치는 영향은 최효미 외(2012)를 참고한다.

사업체 혹은 기업 단위 패널자료를 활용한 표본이탈에 관한 국내의 연구는 많지 않지만, 사업체패널조사를 활용한 몇 개의 연구가 존재한다.

윤윤규 외(2012)는 2007년(2차)과 2009년(3차)의 사업체패널 자료를 활용하여 규모별 이탈 현황을 살펴보았다. 대규모와 소규모 사업체의 이탈은 거의 유사하였는데, 소규모 사업체는 소멸로, 대규모 사업체는 거절로 인한 이탈임을 확인하였다. 이경희 외(2015)는 2011년(4차) 사업체패널 자료를 활용하여 표본이탈의 증가 현황과 사업체의 규모 및 산업의 구조적인 변화로 인한 표본 대표성 훼손을 R-지수(R-indicator)를 활용하여 분석하였다. 분석의 결과 표본 대표성이 훼손되지 않았음을 발견하였다.

이상의 연구를 통해 사업체 혹은 기업 단위 패널조사의 표본이탈 또는 무응답 편향에 대한 분석시 가구나 개인의 패널조사와는 달리 응답자 특성뿐 아니라 응답자가 속해 있는 사업체 특성을 반드시 고려해야 함을 확인하였다. 본 연구는 이러한 선행연구들을 참고하여 우선 표본이탈 현황을 살펴본 후, 표본이탈에 따른 무응답 편향을 분석하여 향후 사업체패널조사의 지속가능성을 제고할 수 있는 방안을 모색해 보고자 한다.

Ⅲ. 자료 소개³⁾

본 연구에서 활용한 자료는 한국노동연구원의 사업체패널 6~9차 조사 자료이고, 여기서는 이 자료를 ‘2015년 표본의 1~4차 조사’로 부른다.

사업체패널조사는 한국의 노동수요와 고용구조 등을 체계적으로 파악하고 전반적인 인적자원 관리체계와 노사관계를 평가하여 바람직한 노동시장 및 노사관계 정책을 수립하기 위한 기초자료로 활용하고자 실시하는 조사이다. 사업체패널조사는 2005년에 5차 조사를 완료한 후 2015년에 신규표본을 설계하고 설문을 일신하여 새롭게 조사를 실시하였다.

사업체패널조사의 설문은 근로자 현황 및 재무 현황, 그리고 사업체 특성, 고용관리, 보상 및 평가, 인적자원관리 및 작업조직, 인적자원개발, 기업복지 및 산업재해, 노사관계 등에 대한 문항으로 구성되어 있다. 사업체패널조사는 사업체의 인력 현황

3) 자세한 사항은 ‘사업체패널 1~9차년도 조사자료 통합유저가이드(2024)’를 참고한다.

을 서면이나 Web을 활용하여 먼저 조사한 다음, 그 외의 설문은 조사원이 패널사업체에 직접 방문하여 컴퓨터를 활용한 대면면접방식인 CAPI(Computer Assisted Personal Interviewing)를 활용하여 조사한다. 이때 재무 현황은 나이스 신용평가 자료의 값으로 대체한다.

사업체패널조사는 전국의 농·임·어업 및 광업 등을 제외한 상시근로자 30인 이상 사업체를 대상으로 하는데, 표본추출틀로 통계청의 ‘전국사업체조사(2014년)’를 활용하여 층화추출하였다. 층은 규모와 산업을 고려하였는데, 규모는 30~99인, 100~299인, 300~499인, 500인 이상의 4가지로, 산업은 경공업, 화학공업, 금속·자동차·운송, 전기·전자·정밀, 건설업, 전기·가스·수도업, 개인서비스업, 유통서비스업, 사업서비스업, 사회서비스업의 10가지로 구분하였다.⁴⁾ 이때 신규표본사업체는 기존 패널모집단의 특성뿐 아니라 표본추가 시점을 기준으로 횡단면모집단의 특성까지 모두 반영할 수 있도록 이중틀(dual frame) 접근방법⁵⁾을 활용하였다. 표본 배분은 예측응답률과 기존 패널표본과의 중복을 함께 고려하여 신규표본의 사업체 수가 약 2,000여 개가 될 수 있도록 다음의 반복절차(iterative procedure)에 의하여 결정하였다.

IV. 무응답 현황 및 편향에 대한 분석결과

1. 응답현황 분석

여기서는 2015년 표본의 2~4차 조사 동안의 응답현황을 응답률 및 거절률, 소멸률(휴·폐업 등) 등을 활용하여 분석하였다. 필요한 경우 과거 사업체패널조사의 2005년 표본 2~4차 조사의 응답현황과 비교하여 살펴보겠다. 이는 국내뿐 아니라 해외에서도 사업체 단위의 패널자료가 많지 않아 2015년 표본의 분석결과와 비교할만한 정보가 적기 때문에 2015년 표본의 품질수준을 상대적으로 파악하기 위함이다.

사업체패널조사의 공공부문은 표본설계시 포함되지 않았으며 개체수도 적고 응

4) 공공부문은 1차 조사에서 20인 이상 기관을 대상으로 별도로 선정하여 조사하여 5차까지 조사된 기관을 대상으로 조사하였다.

5) 박인호·김기민·김정우(2020)를 참고한다.

답률 또한 매우 높아⁶⁾ 분석 데이터셋에서 제외하였다. 즉 민간부문 패널사업체만을 대상으로 분석 데이터셋을 구축하였다.

먼저 2015년 표본의 2~4차 조사 응답현황은 <표 1>과 같다. 여기서 응답률은 식 (1)과 같이 정의하였다.

$$\text{응답률(\%)} = \frac{\text{응답 사업체 수}}{\text{응답 사업체 수} + \text{무응답 사업체 수}} \times 100 \quad (1)$$

<표 1>을 보면, 2015년 표본의 2~4차 조사 응답률은 각각 80.5%, 78.2%, 72.0%로 조사의 차수가 거듭될수록 감소하였다. 패널조사는 보통 패널표본이 구축된 후 실시되는 직후의 조사, 즉 2~3차 조사의 응답률이 크게 떨어진 다음 그 후 조사부터 안정적으로 운영되는 특징을 보인다(Fitzgerald et al. 1998; Lillard et al. 1998, Zabel 1998; 이상호 2005; 최효미 외 2012). 사업체패널조사의 2015년 표본은 2차 조사에서 응답률이 80.5%로 19.5%p 하락한 후 3차 조사부터는 안정적으로 유지되는 것으로 나타났다. 2005년에 실시된 과거 사업체패널의 2~4차 조사의 응답률은 2차 79.9%, 3차 73.2%, 4차 66.5%였고, 국내의 기업 단위 패널조사인 인적 자본기업패널 2~4차 조사의 응답률은 각각 90.3%, 68.3%, 59.7%(황성수 외 2018)였다. 그리고 독일의 IAB Establishment Panel Survey의 1996년 균형 패널표본 2~4차 조사 응답률은 각각 82.0%, 68.2%, 58.8%(Bechmann et al. 2021)로 나타났다.

<표 1> 사업체패널조사 2015년 표본의 응답률 현황 및 비교 자료 (단위: %)

	사업체패널 조사		인적자본 기업패널	독일 IAB establishment 패널조사 (1996년 균형 패널 표본)
	2015년 표본	2005년 표본		
2차	80.5	79.9	90.3	82.0
3차	78.2	73.2	68.3	68.2
4차	72.0	66.5	59.7	58.8

주: 2015년 표본사업체 수는 3,331개소, 2005년 표본사업체 수는 1,615개소, 인적자본기업패널 표본 기업 수는 454개, 독일 IAB establishment 패널조사의 1996년 균형 패널 표본 수는 6,822개임

자료: 저자 작성.

6) 2015년 패널표본 중 공공부문은 100개이며, 공공부문의 응답률은 2차 87%, 3차 97%, 4차 92%임.

사업체패널조사의 2015년 표본의 응답률은 사업체패널조사의 2005년 표본과 인적자본기업패널, 그리고 독일의 IAB Establishment Panel Survey와 비교해 볼 때 대체로 높게 나타나, 사업체패널조사 2015년 표본의 유지율은 상대적으로 양호한 것으로 보인다.

다음으로 응답률과 함께 거절률과 소멸률을 살펴보겠다(<표 2> 참조). 여기서 ‘거절’은 거절, 담당자 부재, 진행중단 등을, ‘소멸’은 휴·폐업, 결번 등을 의미하며, 거절률과 소멸률은 각각 식(2)와 식(3)을 활용하여 계산하였다.

$$\text{거절률(\%)} = \frac{\text{거절 사업체 수}}{\text{응답 사업체 수} + \text{무응답 사업체 수}} \times 100 \quad (2)$$

$$\text{소멸률(\%)} = \frac{\text{소멸 사업체 수}}{\text{응답 사업체 수} + \text{무응답 사업체 수}} \times 100 \quad (3)$$

<표 2> 사업체패널조사 2015년 표본과 2005년 표본의 차수별 응답률, 거절률 및 소멸률
(단위: %)

	2015년 표본 (3,331개소)			2005년 표본 (1,615개소)		
	응답률	거절률	소멸률	응답률	거절률	소멸률
2차	80.5	16.8	2.8	79.9	13.1	7.1
3차	78.2 (2.3)	17.0 (0.2)	4.8 (2.0)	73.2 (6.7)	18.6 (5.5)	8.2 (1.2)
4차	72.0 (6.2)	18.3 (1.3)	9.8 (5.0)	66.5 (6.7)	21.1 (2.5)	12.5 (4.4)
5차				57.7 (8.8)	28.3 (7.2)	14.0 (1.5)

주: 1) 괄호 안은 이전 차수 대비 감소율 혹은 증가율의 차이임.

2) 공공부문은 제외하고 분석함.

자료: 사업체패널 1~9차년도 자료를 활용하여 저자 작성.

2015년 표본의 2~4차 조사 거절률은 각각 16.8%, 17.0%, 18.3%로 증가하는 경향은 보였지만 그 폭은 크지 않았고 2~4차 조사 소멸률은 각각 2.8%, 4.8%, 9.8%로 매 차수마다 약 2배 정도의 크기로 증가하였다.

과거 2005년 표본의 2~5차 조사의 거절률은 각각 13.1%, 18.6%, 21.1%, 28.3%로, 소멸률은 각각 7.1%, 8.2%, 12.5%, 14.0%로 증가하였다. 특히 소멸률은 4차 조사까지 대체적으로 증가폭이 뚜렷하게 커지다가 5차 조사에서 증가폭이 주춤해진 반면, 거절률은 5차 조사에서 7.2%p로 큰 증가폭을 보였다. 2015년 표본의 2~4차 조사의 거절률과 소멸률은 2005년 표본의 2~4차 조사의 거절률과 소멸률에 비해 대체적으로 낮았다.

여기서는 사업체패널조사의 표본설계 기준인 산업과 규모에 따른 응답 현황을 살펴보겠다. 2015년 표본의 규모별 응답률, 거절률 및 소멸률을 분석한 결과는 <표 3>과 같다.

2, 3차 조사에서 500인 이상 사업체의 응답률이 다른 규모 사업체에 비해 2~6%p 정도 높게 나타났으나, 4차 조사에서 500인 이상 사업체의 응답률은 크게 감소하였다. 500인 이상 사업체의 거절률은 2, 3차 조사에서 13.4%로 다른 규모의 사업체에 비해 낮았으나 4차 조사에서 22.5%로 가장 높았다. 소멸률은 조사 차수에 상관없이 가장 작은 규모인 30~99인 사업체에서 가장 높았고, 100~299인 사업체는 4차 조사에서 소멸률의 증가폭이 컸다.

조사가 진행될수록 대규모 사업체는 거절률이 증가하고 중·소규모 사업체는 소멸률이 증가하여, 4차 조사의 응답률은 사업체 규모별로 큰 차이를 보이지 않았다. 이는 윤윤규 외(2012)에서 사업체패널조사의 2005년 표본 2~3차 조사를 분석한 내용과도 유사한 결과이다.

<표 3> 사업체패널조사 2015년 표본의 사업체 규모별 응답률, 거절률 및 소멸률
(단위: 개소, %)

N	응답률			거절률			소멸률			
	2차	3차	4차	2차	3차	4차	2차	3차	4차	
30~99인	1,753	79.6	77.4	71.9	17.3	16.6	16.9	3.1	6.0	11.2
100~299인	873	81.3	78.2	72.1	16.2	18.3	17.9	2.5	3.4	10.1
300~499인	474	80.6	78.7	71.7	17.3	17.9	21.7	2.1	3.4	6.5
500인 이상	231	83.5	83.5	72.7	13.4	13.4	22.5	3.0	3.0	4.8
합계	3,331	80.5	78.2	72.0	16.8	17.0	18.3	2.8	4.8	9.8

자료: 사업체패널 7~9차년도 자료를 활용하여 저자 작성.

<표 4> 2015년 표본의 산업별 응답률, 거절률 및 소멸률 현황 (단위: 개소, %)

	N	응답률			거절률			소멸률		
		2차	3차	4차	2차	3차	4차	2차	3차	4차
경공업	438	80.8	78.8	71.9	15.8	15.8	18.0	3.4	5.5	10.0
화학공업	301	83.4	82.1	74.1	15.6	15.3	21.3	1.0	2.7	4.7
금속·자동차·운송	452	84.5	80.8	73.9	12.2	14.8	17.7	3.3	4.4	8.4
전기·전자·정밀	269	79.2	79.2	73.6	17.5	16.0	17.8	3.3	4.8	8.6
건설업	272	83.5	73.5	70.6	13.6	21.3	15.4	2.9	5.1	14.0
개인서비스업	414	73.9	73.4	66.7	22.2	18.8	19.8	3.9	7.7	13.5
유통서비스업	312	82.7	75.6	69.9	15.4	20.8	20.8	1.9	3.5	9.3
사업서비스업	548	75.0	75.9	69.7	22.4	19.2	17.9	2.6	4.9	12.4
사회서비스업	293	86.0	86.3	80.5	11.6	10.9	14.7	2.4	2.7	4.8
전기·가스·수도업	32	81.3	81.3	71.9	18.8	12.5	21.9	0.0	6.3	6.3
합계	3,331	80.5	78.2	72.0	16.8	17.0	18.3	2.8	4.8	9.8

자료: 사업체패널 7~9차년도 자료를 활용하여 저자 작성.

2015년 표본의 산업별 응답률, 거절률 및 소멸률을 분석한 결과는 <표 4>와 같다. 개인서비스업과 사업서비스업 사업체는 2~4차 조사 동안 대체로 낮은 응답률을 보였고, 건설업과 유통서비스업 사업체는 2차 조사 이후 낮은 응답률을 보였다. 반면, 사회서비스업 사업체는 2~4차 조사 동안 높은 응답률을 유지하였다. 개인서비스업과 사업서비스업 사업체는 2~3차 조사에서는 거절률이, 4차 조사에서는 소멸률이 높았고, 건설업 사업체는 3차 조사에서는 거절률이, 4차 조사에서는 소멸률이 높았으며, 유통서비스업 사업체는 3, 4차 조사 모두 거절률이 높게 나타나 전반적으로 낮은 응답률을 보였다.

2. 무응답 편향

Petroni et al. (2004)는 사업체 조사의 무응답 편향은 모집단의 왜곡을 크게 야기할 수 있어 가구 조사에 비해 더 큰 문제가 될 수 있음을 강조하였다. 여기서는 2015년 표본의 2~4차 조사의 무응답 편향을 살펴보겠다. 무응답 편향은 3가지 분

석방법을 활용하였다. 첫째, König et al.(2021)의 분석방법을 참고하여 주요 변수에 대한 무응답 및 거절, 비접촉 편향을 비교 가능하도록 계산하였다. 둘째, 하위집단(subgroup) 응답률의 변동계수를, 마지막으로 무응답 가중치의 분산을 분석하였다.

먼저 주요 변수에 대한 무응답 및 거절, 비접촉 편향을 살펴보겠다. 1차 조사 자료에는 응답한 사업체뿐 아니라 이탈된 사업체의 응답도 포함되어 있어, 무응답 편향을 측정하기 위해 1차 조사의 값을 활용하여 분석하였다.

무응답 편향(nonresponse bias)을 측정하기 위해 표본설계 변수 14개, 사업체 특성 관련 변수 20개, 노동력 구성 변수 9개 등 사업체패널 조사의 주요 변수를 선택하였다.⁷⁾ 거절 편향(refusal bias)과 비접촉 편향(non-contact bias) 또한 동일한 변수를 사용하였다. 여기서 ‘무응답’은 ‘거절’과 ‘비접촉’인 경우를 의미하고, ‘거절’은 단순 거절, 담당자 부재, 진행중단 등을, ‘비접촉’은 휴·폐업, 결번 등을 말한다. 무응답 편향(\overline{Y}_{nr}), 거절 편향(\overline{Y}_{rf}), 비접촉 편향(\overline{Y}_{nc})의 추정식은 각각 식(4)~(6)과 같다(Konig et al. 2021; Sakshaug et al. 2019; Sakshaug & Huber 2016; Groves 2006).

$$\overline{Y}_{nr} = \overline{Y}_r - \overline{Y}_t \tag{4}$$

$$\overline{Y}_{rf} = \overline{Y}_r - \overline{Y}_c \tag{5}$$

$$\overline{Y}_{nc} = \overline{Y}_c - \overline{Y}_t \tag{6}$$

여기서 \overline{Y}_t 는 전체 사업체에 대한 해당 변수의 평균, \overline{Y}_r 는 응답한 사업체에 대한 해당 변수의 평균, \overline{Y}_c 는 접촉한 사업체(응답 사업체+거절 사업체)에 대한 해당 변수의 평균을 의미한다.

추정된 편향의 크기는 서로 다른 범주의 변수들을 함께 해석하기 어렵기 때문에 절대 상대 무응답 편향(absolute relative nonresponse bias)을 식(7)~(9)와 같이 추정하였다. 거절 편향과 비접촉 편향 또한 동일한 방식으로 추정하였다(Konig et al. 2021; Sakshaug et al. 2019; Sakshaug & Huber 2016; Groves 2006).

7) 여기서 활용한 변수는 사업체패널 조사의 주요 변수이므로 향후 응답 성향 모형에서도 동일하게 활용하였다. 변수에 대한 자세한 정보는 응답 성향 모형을 참고한다.

$$\widetilde{Y}_{nr} = \left| \frac{\overline{Y}_r - \overline{Y}_t}{\overline{Y}_t} \right| \tag{7}$$

$$\widetilde{Y}_{rf} = \left| \frac{\overline{Y}_r - \overline{Y}_c}{\overline{Y}_c} \right| \tag{8}$$

$$\widetilde{Y}_{nc} = \left| \frac{\overline{Y}_c - \overline{Y}_t}{\overline{Y}_t} \right| \tag{9}$$

여기서 \widetilde{Y}_{nr} , \widetilde{Y}_{rf} , \widetilde{Y}_{nc} 는 각각 절대 상대 무응답, 거절 및 비접촉 편향이고 \overline{Y}_t , \overline{Y}_r 과 \overline{Y}_c 는 식(5)~(7)과 같다.

결과를 요약하기 위해 절대 상대 무응답, 거절 및 비접촉 편향의 평균을 구하였고, 그 결과는 <표 5>와 같다.⁸⁾

<표 5> 2015년 표본의 절대 상대 무응답 및 거절, 비접촉 편향 (단위: %)

	무응답 편향				거절 편향				비접촉 편향			
	전체	표본 설계	사업체 특성	노동력 구성	전체	표본 설계	사업체 특성	노동력 구성	전체	표본 설계	사업체 특성	노동력 구성
2차	2.26	3.20	2.12	1.12	2.01	2.83	1.76	1.28	0.58	0.72	0.61	0.30
3차	3.09	4.26	3.18	1.10	2.32	3.56	3.18	0.08	1.10	1.32	1.19	0.57
4차	2.68	2.67	3.15	1.62	1.65	2.10	1.75	0.73	2.37	2.99	2.28	1.59

자료: 사업체패널 6~9차년도 자료를 활용하여 저자 작성.

2015년 표본의 2~4차 조사의 무응답 편향은 각각 2.26%, 3.09%, 2.68%로, 3차 조사까지 증가하다 4차 조사에서 다소 감소하였다. 거절 편향은 2차 2.01%, 3차 2.32%, 4차 1.65%로, 3차까지 증가하다 4차에서 감소하였는데, 4차 조사의 거절 편향이 가장 작았다. 비접촉 편향은 2차 0.58%, 3차 1.10%, 4차 2.37%로 증가하였다. 편향의 크기를 판단하는 절대적인 기준은 없지만 Sakshaug et al.(2019)는 실질적인 임계값을 10%라고 제시하였다(König et al. 2021). 2015년 표본의 무응답 편향은

8) 세부 변수에 대한 자세한 편향은 이경희 외(2023)의 부록을 참고한다.

2~3% 수준으로 Sakshaug et al.(2019)가 채택한 기준인 10%에 비해 작게 나타나 무응답 편향이 크지 않음을 확인하였다. 거절 편향은 증감을 보이기도 했지만 비접촉 편향은 증가되는 방향으로 나타났다. 거절 편향과 비접촉 편향은 표본설계 변수에서 가장 큰 편향을 보였다.

다음으로 하위집단 응답률의 변동계수를 분석하였다. 하위집단 응답률의 변동계수는 하위집단 응답률을 요약한 척도로서, 표본설계 변수를 이용하여 각 셀을 만들어 셀별 응답률을 활용하여 계산하였다. 하위집단 응답률의 변동계수는 각 셀(cell)별 응답률 차이가 클수록 무응답 편향이 커질 수 있기 때문에 이 값의 변동이 클수록 대표성은 떨어진다고 볼 수 있다(손창균 2015). 하위집단 응답률의 변동계수는 식(10)과 같이 계산하였다. 여기서 K 는 하위집단 수, R 은 전체 표본의 응답률, R_k 는 하위집단의 응답률이다.

$$CV = \frac{\sqrt{\frac{1}{K-1} \sum_{k=1}^K (R_k - R)^2}}{R} \tag{10}$$

마지막으로 무응답 가중치의 분산을 구하였다. 무응답 가중치의 분산은 무응답 조정 셀에 무응답을 조정하기 위해 부여된 가중치의 분산을 의미한다. 이 값 역시 커지면 무응답 편향의 위험도가 증가하게 되므로 대표성이 떨어지게 된다(손창균 2015). 무응답 가중치의 분산은 식(11)과 같이 계산하였다. 이때 $w_{nr,k}$ 는 응답률의 역수($= R_k^{-1}$, R_k 는 각 셀별 응답률), n_r 는 응답 사업체 수, K 는 무응답 조정 셀, n_k 는 각 셀별 응답 사업체 수이다.

$$Var(w_{nr}) = \frac{1}{n_r - 1} \sum_{k=1}^K \sum_{i=1}^{n_k} (w_{nr,k_i} - \bar{w}_{nr})^2 \tag{11}$$

여기서 $\bar{w}_{nr} = \frac{1}{n_r} \sum_{k=1}^K \sum_{i=1}^{n_k} w_{nr,k_i}$ 이다.

<표 6> 2015년 표본의 하위집단 응답률의 변동계수와 무응답 가중치의 분산

	하위집단 응답률의 변동계수(%)	무응답 가중치의 분산
2차	7.9	0.006
3차	9.2	0.008
4차	9.9	0.010

자료: 사업체패널 6~9차년도 자료를 활용하여 저자 작성.

<표 6>은 하위집단 응답률의 변동계수와 무응답 가중치의 분산을 분석한 결과이다. 하위집단 응답률의 변동계수는 2차 조사 7.9%, 3차 조사 9.2%, 4차 조사 9.9%로, 무응답 가중치의 분산은 2차 조사 0.006, 3차 조사 0.008, 4차 조사 0.010으로 두 경우 모두 증가하였다. 하위집단 응답률의 변동계수는 4차 조사까지 10% 미만으로 나타났고, 무응답 가중치의 분산 또한 크게 증가하지 않아 2015년 표본의 셀별 응답률은 상당히 안정적인 것으로 보인다.⁹⁾

3. 응답 성향

2015년 표본의 2~4차 조사 자료를 결합(pooled)하여 응답 성향을 살펴보았다. 2015년 표본의 1차 조사 자료는 응답 사업체뿐 아니라 이탈된 사업체의 응답도 포함되어 있어, 응답 사업체와 이탈된 사업체의 특성 차이가 있는지 1차 조사의 값을 활용하여 분석하였다. 응답 성향 모형은 종속변수가 이진변수일 때 흔히 사용하는 probit을 활용하였고, 모형식은 식(12)와 같다. 이때 설명변수는 표본설계시 고려했던 변수와 사업체패널조사의 사업체 특성과 노동력 구성을 파악할 수 있는 주요 변수들로 구성하였다.

$$y_i^* = \alpha + \beta x_i + e_i, \quad y_i = 1[y_i^* = 1] \quad (12)$$

여기서 y_i^* 는 관측되지 않는 잠재변수(latent variable)이고 관측된 종속변수 y_i 는 응답여부(응답=1, 무응답=0)를 나타낸다. 설명변수 x_i 는 표본설계 및 사업체 특성, 노동력 구성 변수로 구성하였다. 표본설계 변수는 규모(99인 이하, 100~299인, 300~

9) Kish(1965, p218)에 의하면 변동계수가 10% 미만이면 충분히 양호(sufficiently good)하고 20% 미만이면 허용할 수 있다(tolerable)라고 했다.

499인, 500인 이상), 산업(경공업, 화학공업, 금속·자동차·운송, 전기·전자·정밀, 건설업, 개인서비스업, 유통서비스업, 사업서비스업, 사회서비스업, 전기·가스·수도업), 사업체 특성 변수는 사업체 업력, 단독사업체, 해외진출, 외국인 지분율, 정규직 해고, 유노조, 임금피크제 실시, 인사평가 실시, 성과배분제 있음, 정년제도 있음, 공정한 기회 관련 정책 있음, 고충처리위원회 있음, 교육훈련 실시, 육아휴직제도 있음, 경쟁 심함, 가격 저렴, 품질 좋음, 업무로테이션 실시, 다기능 교육훈련 실시, 업무제한제도 실시이고, 노동력 특성 변수는 여성 비율, 55세 이상 비율, 35세 미만 비율, 이직률, 비정규직 비율, 관리직 비율, 전문·사무직 비율, 서비스·판매직 비율, 생산·단순직 비율 등이다. e_i 는 통상적인 오차항이다.

응답 성향 모형의 분석결과는 <표 7>과 같다. 응답 성향 모형의 설명력은 0.030으로 매우 낮았다. Fitzgerald et al.(1998)과 이상호(2005)에 나타난 표본이탈 모형의 Pseudo R^2 또한 각각 0.028~0.071, 0.036으로 낮았고, König et al.(2021)는 로지스틱 모형의 Pseudo R^2 를 제시하지 않았지만 변수의 한계효과가 대부분 작고 ‘0’에 가까운 값이었다. 이는 일반적인 표본이탈 모형에서 조사원 효과, 응답자의 이탈 직전 현황(개인인 경우 이혼, 실업, 질병 등, 사업체인 경우 합병, 매각 등)과 같이 관측이 쉽지 않거나 수치화가 어려운 요인이 설명변수에서 누락되어 나타난 현상일 것으로 판단된다(이상호, 2005). 향후 정교한 응답 성향 모형을 위해 paradata 등을 활용하여 조사과정을 수치화하거나 외부자료를 통해 사업체의 경제현황을 판단할 수 있는 변수작성에 대한 고민이 필요할 것으로 보인다.

이제 probit 분석결과를 변수별로 자세히 살펴보겠다. 먼저 표본설계에 활용된 규모와 산업 변수를 살펴보면, 통계적으로 유의하지는 않았지만 큰 규모 사업체는 작은 규모 사업체에 비해 응답 확률이 낮게 나타났고, 이는 다른 연구결과와 일치한다(Tomaskovic-Devey, Leiter, & Thompson 1994; Earp et al. 2018; König et al. 2021, König & Sakshaug 2023) 경공업 사업체에 비해 금속·자동차·운송 제조업 사업체와 전기·가스·수도업 사업체는 응답 확률이 높고 유통서비스업과 사업서비스업 사업체는 응답 확률이 낮은 것으로 통계적으로 유의하게 나타났다.

다음으로 사업체 특성 중 통계적으로 유의한 변수를 중심으로 살펴보면, 업력이 긴 사업체일수록 응답 확률이 높고 해외진출한 사업체일수록 응답 확률이 낮았다. 임금피크제를 실시하는 사업체는 그렇지 않은 사업체에 비해, 고충처리위원회가 있는 사업체는 그렇지 않은 사업체에 비해, 교육훈련을 실시하는 사업체는 그렇지 않

은 사업체에 비해, 육아휴직제도가 있는 사업체는 그렇지 않은 사업체에 비해 응답 확률이 높았다. 국내 시장에서의 경쟁이 심하다고 응답한 사업체와 경쟁업체에 비해 품질이 좋다고 응답한 사업체는 그렇지 않은 사업체에 비해 응답확률이 낮았다. 이를 통해 사업체의 제도적 혹은 경영적 측면에 따라 응답 성향이 다를 수 있음을 확인하였다.

노동력 구성 변수 중 중 통계적으로 유의한 변수를 중심으로 살펴보면, 여성 근로자의 비율이 높을수록 응답확률이 낮고 55세 이상 고령 근로자의 비율이 높을수록 응답확률이 높았다

2차 조사에 비해 3차 조사의 응답확률은 2.2%p, 4차 조사의 응답확률은 8.5%p 낮게 나타났다.

<표 7> 2015년 패널표본의 응답 확률(probit)

		계수	한계효과	표준오차
[기준: 30~99인]				
	100~299인 이하=1	0.009	0.003	(0.036)
	300~499인=1	-0.080	-0.024	(0.051)
	500인 이상=1	-0.004	-0.001	(0.061)
[기준: 경공업]				
표본 설계	화학공업=1	0.059	0.017	(0.065)
	금속·자동차·운송=1	0.130	0.038	(0.059)**
	전기·전자·정밀=1	0.067	0.020	(0.065)
	건설업=1	-0.051	-0.015	(0.074)
	개인서비스업	-0.003	-0.001	(0.153)
	유통서비스업=1	-0.117	-0.035	(0.062)*
	사업서비스업=1	-0.222	-0.066	(0.076)***
	사회서비스업=1	-0.083	-0.025	(0.060)
	전기·가스·수도업=1	0.259	0.076	(0.082)***
사업체 특성	사업체 업력(年)	0.008	0.002	(0.001)***
	단독사업체=1	-0.055	-0.016	(0.034)
	해외진출합=1	-0.193	-0.057	(0.042)***
	외국인 지분율(%)	-0.001	0.000	(0.001)

사업체 특성	정규직 해고=1	0.090	0.027	(0.056)
	유노조 사업체=1	0.058	0.017	(0.044)
	임금피크제 실시=1	0.102	0.030	(0.049)**
	인사평가 실시=1	0.030	0.009	(0.034)
	성과배분제 있음=1	-0.046	-0.014	(0.037)
	정년제도 있음=1	-0.020	-0.006	(0.035)
	공정한 기회 문서화=1	0.043	0.013	(0.039)
	고충처리위원회 있음=1	0.062	0.018	(0.036)*
	교육훈련 실시=1	0.120	0.036	(0.032)***
	육아휴직제도 있음=1	0.086	0.026	(0.042)**
	경쟁이 심함=1	-0.119	-0.035	(0.032)***
	가격이 저렴=1	0.002	0.000	(0.044)
	품질이 좋음=1	-0.051	-0.015	(0.030)*
	업무로테이션 실시=1	-0.006	-0.002	(0.045)
	다기능 교육훈련 실시=1	-0.030	-0.009	(0.037)
업무제한제도 실시=1	0.020	0.006	(0.034)	
노동력 구성	여성비율	-0.166	-0.049	(0.071)**
	55세 이상 비율	0.149	0.044	(0.085)*
	35세 미만 비율	0.104	0.031	(0.078)
	이직률	0.105	0.031	(0.066)
	비정규직 비율	0.024	0.007	(0.065)
	전문·사무직 비율	0.069	0.020	(0.137)
	서비스·판매직 비율	0.050	0.015	(0.133)
	생산·단순직	-0.005	-0.002	(0.129)
차수	[기준: 2차 조사]			
	3차 조사=1	-0.080	-0.022	(0.035)**
	4차 조사=1	-0.283	-0.085	(0.034)***
<i>N</i>		9,993		
<i>R</i> ²		0.030		
Log-likelihood		-5,242.86		

주: *, **, ***는 각각 유의수준 10%, 5%, 1%에서 유의함.

자료: 사업체패널 6-9차년도 자료를 활용하여 저자 계산.

V. 모의실험

본 장에서는 Monte Carlo Simulation을 활용하여 향후 실시할 2015년 표본의 5차 조사 응답 상황을 예측해 보았다. 이때 사업체의 무응답 편향은 모집단의 왜곡을 크게 발생시킬 수 있으므로 제4장에서 제시한 세 가지 무응답 편향 값을 구하여 살펴보았다.

모의실험은 다음과 같이 설정하였다. 먼저, 4차 조사를 기준으로 산업과 규모 셀별 응답률을 계산하여 개체에 반영한다. 이때 4차 조사에서 거절 및 소멸인 경우, 동일하게 셀별 응답률을 반영하게 되면 예상 응답률이 높아지기 때문에 3차 조사에서 거절 및 소멸한 경우의 응답률을 4차 조사 셀별 응답률로 대체하여 활용했다. 다음으로, uniform 분포로 난수를 발생시켜 앞에 계산한 응답률보다 작은 경우는 5차 조사 응답(1)으로, 큰 경우는 5차 조사에서 무응답(0)으로 처리한다. 그리고 응답률을 셀별로 동일 비율로 감소시키면서 응답률, 무응답 편향, 하위집단 응답률의 변동계수, 무응답 가중치의 분산을 계산한다. 이 과정을 1,000번 반복했다.

2015년 표본을 가지고 모의실험하여 5차 조사를 예측하였다. 예측결과를 평가하기 위해 2005년 표본에도 동일하게 적용하여 실제 5차 조사의 결과와 비교하였다. 2015년 표본의 모의실험 결과는 <표 8>이고 2005년 표본의 모의실험 결과는 <표 9>이며, 이를 도식화한 결과는 <그림 1>~<그림 3>이다. 2005년 표본을 분석한 <표 9>를 보면, 실제 5차 조사의 응답률인 57.7%(<표 2> 참조)와 유사한 응답률 57.8%에서 무응답 편향은 6.15, 하위집단 응답률의 변동계수는 22.9, 무응답 가중치의 분산은 0.133으로 나타나, 실제 무응답 편향 5.98, 하위집단 응답률의 변동계수 22.9, 무응답 가중치의 분산 0.115와 유사함을 확인하였다. 따라서 설정한 모의실험이 사업체패널조사의 향후 상황을 예측하는 데 적절한 것으로 판단된다.

2015년 표본의 사업체패널조사는 조사대상 제외 사업체를 뺀 유효표본 대비 75~80% 패널유지를 목표로 하고 있어, 분석대상 사업체 기준으로 5차 조사는 약 2,000~2,200여 개, 6차 조사는 1,700~2,000여 개 사업체에 대해 조사를 완료할 것으로 예상된다. 5, 6차 조사가 완료될 것으로 예상되는 사업체 수에 대한 응답률 및 응답 편향에 대해 예측해보면(<표 8> 참조), 6차 조사의 무응답 편향은 3.5~4.0%이

고 변동계수는 16~19%로 용인가능한 수준으로 나타나, 향후 6차 조사까지 시행할 수 있을 것으로 보인다.

<표 8> 2015년 표본의 절대 상대 무응답 및 거절, 비접촉 편향 (단위: 개소, %)

응답률	응답 개체 수	무응답 편향	하위집단 응답률의 변동계수	무응답 가중치의 분산
66.3	2,207	3.35	15.1	0.031
64.6	2,150	3.42	15.7	0.034
62.7	2,090	3.52	16.0	0.037
61.0	2,031	3.54	16.5	0.041
59.2	1,973	3.64	16.9	0.046
57.4	1,913	3.71	17.5	0.052
55.5	1,850	3.77	18.1	0.058
53.8	1,793	3.84	18.4	0.066
52.0	1,732	3.93	19.3	0.076
50.2	1,672	4.07	19.7	0.084
48.4	1,613	4.17	20.5	0.099
46.5	1,550	4.20	21.0	0.111
44.8	1,493	4.36	22.0	0.129
43.1	1,435	4.42	22.5	0.151
41.2	1,371	4.57	23.2	0.173
39.3	1,309	4.73	24.2	0.206
37.7	1,255	4.81	24.9	0.243
35.8	1,192	4.96	25.9	0.290
34.1	1,137	5.13	26.7	0.336
32.2	1,072	5.22	27.9	0.407
30.5	1,015	5.48	29.1	0.499

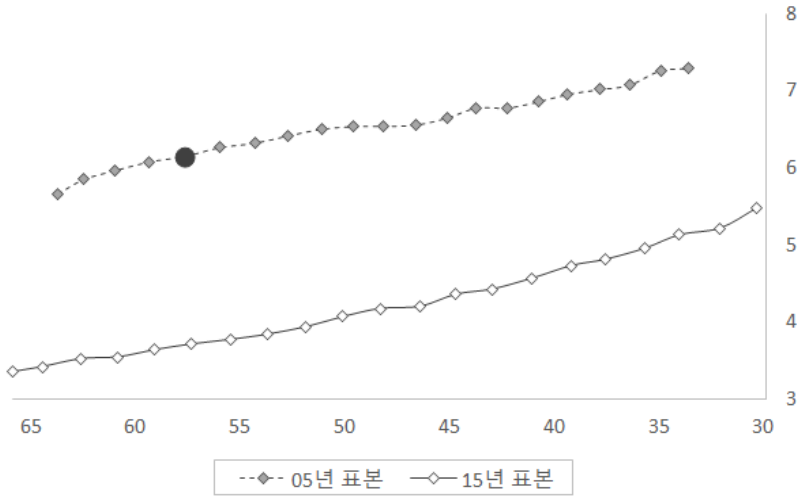
자료: 사업체패널 6~9차년도 자료를 활용하여 저자 작성.

<표 9> 2005년 표본의 모의실험 결과

(단위: 개소, %)

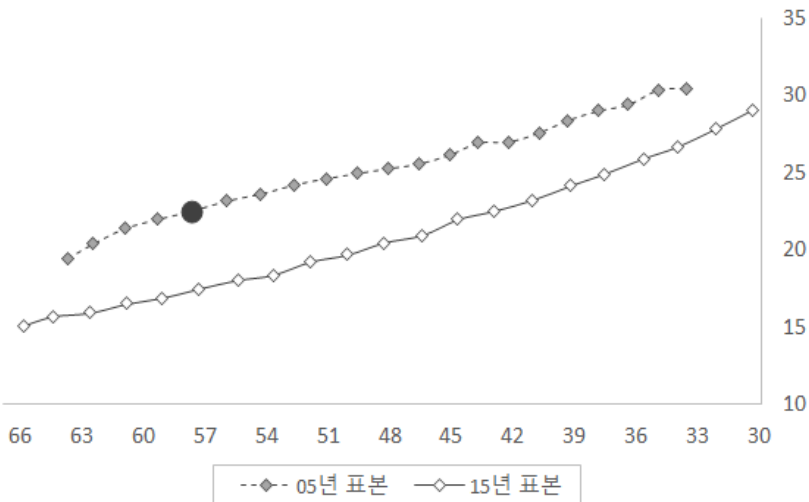
응답률	응답 개체 수	무응답 편향	하위집단 응답률의 변동계수	무응답 가중치의 분산
63.8	1,031	5.66	19.5	0.084
62.6	1,011	5.85	20.4	0.096
61.1	986	5.97	21.4	0.108
59.5	960	6.08	22.0	0.120
57.8	933	6.15	22.5	0.133
56.1	906	6.27	23.2	0.147
54.4	879	6.32	23.6	0.161
52.8	853	6.42	24.2	0.179
51.2	827	6.51	24.6	0.195
49.7	803	6.54	25.0	0.210
48.2	779	6.54	25.3	0.227
46.7	755	6.56	25.6	0.246
45.2	730	6.65	26.2	0.272
43.8	708	6.78	26.9	0.298
42.3	684	6.78	27.0	0.320
40.9	660	6.86	27.6	0.360
39.5	637	6.96	28.4	0.405
37.9	613	7.02	29.0	0.442
36.5	590	7.08	29.4	0.496
35.0	566	7.27	30.4	0.567
33.7	544	7.29	30.4	0.624

자료: 사업체패널 1~5차년도 자료를 활용하여 저자 작성.



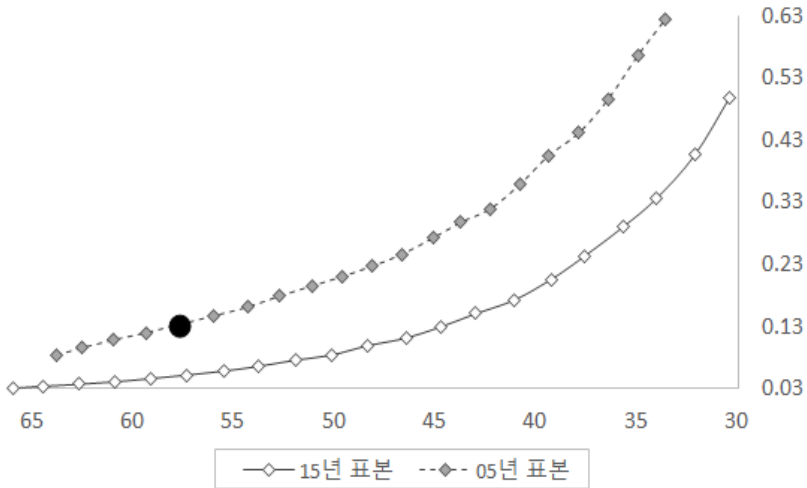
<그림 1> 무응답 편향에 대한 모의실험 결과

- 주: 1) 가로축은 응답률, 세로축은 무응답 편향 값임.
 2) 검정색 원은 2005년 표본의 5차 조사 결과와 가장 유사한 경우를 표시한 것임.
 자료: 사업체패널 1~9차년도 자료를 활용하여 저자 작성.



<그림 2> 하위집단 응답률의 변동계수에 대한 모의실험 결과

- 주: 1) 가로축은 응답률, 세로축은 응답률 변동계수 값임.
 2) 검정색 원은 2005년 표본의 5차 조사 결과와 가장 유사한 경우를 표시한 것임.
 자료: 사업체패널 1~9차년도 자료를 활용하여 저자 작성.



<그림 3> 무응답 가중치의 분산에 대한 모의실험 결과

주: 1) 가로축은 응답률, 세로축은 무응답 가중치의 분산 값임.
 2) 검정색 원은 2005년 표본의 5차 조사 결과와 가장 유사한 경우를 표시한 것임.
 자료: 사업체패널 1~9차년도 자료를 활용하여 저자 작성.

지금까지의 분석결과를 통해 볼 때 2015년 표본의 2~4차 조사는 응답률도 높고 응답 편향도 작아 상당히 양호한 수준인 것으로 나타났다. 또한 향후에도 2차례 정도의 패널조사를 실시할 수 있을 것으로 예측되었다.

VI. 요약 및 결론

본 연구는 2015년에 새롭게 추출한 사업체패널조사 패널표본을 활용하여 표본 이탈 현황을 분석하고 표본이탈이 대표성과 추정치 편향에 미치는 영향을 분석해 봄으로써 향후 사업체패널 조사의 지속가능성을 검토하였다.

2015년 표본의 응답률은 2차 조사 80.5%, 3차 조사 78.5%, 4차 조사 72.0%로 과거 2005년 표본의 응답률(2차 79.9%, 3차 73.2%, 4차 66.5%)과 국내·외 다른 조사에 비해서도 높아 2015년 표본 유지율은 상대적으로 양호하였다. 2015년 표본

의 무응답 편향은 2차 조사 2.26%, 3차 조사 3.09%, 4차 조사 2.68%로 크지 않았다. 또한 무응답 편향을 검정할 수 있는 하위집단 응답률의 변동계수는 2~4차 조사 동안 10% 미만(2차 7.9%, 3차 9.2%, 4차 9.9%)을 유지했고 무응답 가중치의 분산 또한 2차 0.006, 3차 0.008, 4차 0.010으로 크게 증가하지 않았다. 이를 통해 2015년 표본은 4차례의 조사가 진행되는 동안 패널표본의 대표성이 크게 훼손되지 않았음을 확인하였다. 응답 성향 모형을 통해, 유의하지는 않았으나 대규모 사업체가 소규모 사업체에 비해 응답확률이 낮음과 사업체의 제도적 혹은 경영적 측면에 따라 응답 가능성이 다름을 확인하였다.

Monte Carlo Simulation을 활용하여 향후 실시될 2015년 표본의 조사응답 상황을 예측해 본 결과, 사업체패널조사는 향후 2차례 정도 조사를 하여도 데이터의 품질(quality)을 잘 관리할 수 있을 것으로 나타났다.

종합하면, 사업체패널조사는 2015년 신규표본을 구축한 이후 4차 조사까지 잘 유지되고 있고 향후 2차례 정도의 조사를 실시할 수 있음을 확인하였다. 그러나 4차 조사에서 500인 이상 대규모 사업체 중 3개 산업의 응답 사업체 수가 10개 이하로 매우 적어 해당 셀의 응답률이 낮아진다면 5, 6차 조사에서 무응답 편향이 현재의 예측보다 커질 수 있다. 이를 예방하기 위해서는 대규모 사업체 등과 같이 응답 확률이 낮은 특성을 보이는 사업체의 참여를 높일 수 있는 방안을 모색할 필요가 있다.

본 연구의 응답 성향 모형은 사업체 특성과 노동력 구성 등에만 초점을 맞춰 분석하였는데, 추후에는 조사기간, 응답시간, 조사원과 응답자의 교체, 조사원의 숙련도 등과 같은 조사환경적 요인 또한 고려한 연구가 필요할 것이다.

참고문헌

- 박인호·김기민·김정우. 2020. “이중틀 접근을 통한 사업체패널조사의 표본추가 및 가중치 산출.” 《조사연구》 21(3): 37-59.
- 손창균. 2015. 《2015년 여성가족패널조사 표본보완을 위한 표본설계 최종보고서》 한국여성정책연구원.
- 윤윤규·홍민기·이상호·배기준·최효미·김기민·장숙량·조성일·천희란. 2012. 《패널자료 품질개선 연구(Ⅱ)》 한국노동연구원.

- 이경희·박인호·이상은·신기일·한치록·이고은. 2015. 《패널자료 품질개선 연구(V)》 한국노동연구원.
- 이경희·정현상·권익성·김정우·김기민·황선웅·Zhexun Mo·Li Yang. 2023. 《패널자료 품질개선 연구(XIII)》 한국노동연구원.
- 이상호. 2005. “한국노동패널(KLIPS)의 표본이탈 분석-가구소득을 중심으로.” 《노동리뷰》 11: 66-80.
- 최효미·이상호·성재민·김기민·배기준. 2012. 《패널자료의 품질 개선을 위한 조사기법 연구 - 컴퓨터를 이용한 조사(CAPI) 방법-》 한국노동연구원.
- 황성수·민주홍·윤혜준·황승록·박라인. 2018. 《인적자본기업패널조사(2018)》, 한국직업능력연구원.
- 한국노동연구원. 2024. 《사업체패널 1~9차년도 조사자료 통합유저가이드(version 1.81)》 <https://www.kli.re.kr/wps>
- Bechmann, S., N. Tschersich, P. Ellguth, S. Kohaut, and C. Florian. 2021. “Technical Report on the IAB Establishment Panel-Wave28(2020).” in FDZ-Methodenreport 07|2021, Bundesagentur für Arbeit: Nurnberg.
- Earp, M., D. Toth P. Phipps, and C. Oslund. 2018. “Assessing Nonresponse in a Longitudinal Establishment Survey Using Regression Trees.” *Journal of Official Statistics* 34(2): 463-481.
- Fitzgerald, J., P. Gottschalk, and R. Moffitt. 1998. *An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics* 33(2): 251-299.
- Groves, R.M. 2006. “Nonresponse Rates and Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70(5): 646-675.
- Hidiroglou, M.A., J.D. Drew, and G.B. Gray. 1993. “A Framework for Measuring and Reducing Nonresponse in Surveys.” *Survey Methodology* v. 19(1):81-94.
- Janik, F. and S. Kohaut. 2012. “Why Don’t They Answer? Unit Nonresponse in the IAB Establishment Panel.” *Quality & Quantity* 46: 917-934.
- Lillard, Lee A. and W.A. Panis Constantijn. 1998. “Panel Attrition from the Panel Study of Income Dynamics: Household Income, Marital Status, and Mortality.” *The Journal of Human Resources* 32(2): 437-457.
- Kish, L. 1965. *Survey Sampling*. John Wiley and Sons, Inc., New York.
- König, C. and J.W. Sakshaug. 2023. “Nonresponse Trends in Establishment Panel Surveys: Findings from the 2001-2017 IAB Establishment Panel.” *Journal of Labor Market Research Springer; Institute for Employment Research/Institut für Arbeitsmarkt-und Berufsforschung (IAB)* 57(1): 1-17, December.

- König, C., J.W. Sakshaug, J. Stegmaier, and S. Kohaut. 2021. “Trends in Establishment Survey Nonresponse Rates and Nonresponse Bias: Evidence from the 2001-2017 IAB Establishment Panel.” *Journal of Official Statistics* 37(4): 931-953.
- Petroni, R., R. Sigman, D. Willimack, S. Cohen, and C. Tucker. 2004. “Response Rates and Nonresponse in Establishment Survey-BLS and Census Bureau.” *Federal Economic Statistics Advisory Committee*: 1-50
- Sakshaug, J.W. and M. Huber. 2016 “An Evaluation of Panel Nonresponse and Linkage Consent Bias in a Survey of Employees in Germany.” *Journal of Survey Statistics and Methodology* 4(1): 71~93.
- Sakshaug, J.W., S. Hulle, A. Schmucker, and S. Liebig. 2019 “Panel Survey Recruitment With or Without Interviewers?—Implications for Nonresponse Bias, Panel Consent Bias, and Total Recruitment Bias.” *Journal of Survey Statistics and Methodology* 8(3): 540~565.
- Tomaskovic-Devey, D., J. Leiter, and S. Thompson. 1994. “Organizational Survey Response.” *Administrative Science Quarterly* 39(3): 439-457.
- Zabel, J.E. 1998. “An Analysis of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an Application to a Model of Labor Market Behavior.” *The Journal of Human Resources* 33(2): 479-506

<접수 2024.02.14; 수정 2024.03.08; 게재확정 2024.04.03>

Review of Sample Attrition and Nonresponse Bias of Workplace Panel Survey

Kim, Kimin
(Korea Labor Institute)

This study examines the sustainability of future Workplace Panel Surveys(WPS) by analyzing the status of sample attrition and the impact of attrition on representative of sample and estimation bias using a new sample of WPS in 2015.

The response rate of the 2015 WPS was 80.5% in Wave 2, 78.5% in Wave 3, and 72.0% in Wave 4, which was generally higher than the response rate of the 2005 WPS and higher than other surveys in Korea and abroad. The nonresponse bias in the 2015 WPS was not significant, at 2.26% in Wave 2, 3.09% in Wave 3, and 2.68% in Wave 4. The coefficient of variation of the sub-group response rate, which is a test for nonresponse bias, remained below 10% from Wave 2 to Wave 4, and the variance of the nonresponse weights did not increase significantly. Through this, it was confirmed that the representativeness of the 2015 WPS was not significantly compromised over the four waves of the survey. Using Monte Carlo Simulations to predict the response situation of the 2015 WPS sample in future surveys, we found that the data quality of the WPS can be well maintained for as few as two future surveys.

This study confirms that the WPS has been well maintained since the new sample was established in 2015 until the fourth survey. Considering the results of the four waves and the simulation results, it is likely that the survey can be conducted twice in the future, but if the nonresponse rate increases in certain groups in the future, the results may differ from the current forecast. In order to prevent this, it is worth exploring ways to increase engagement with workplace that have been shown to be less likely to respond, such as large-scale workplace.

Key words: Workplace Panel Survey, attrition, response rate, non-response bias, Monte Carlo simulation