

연구논문

이중틀 접근을 통한 사업체패널조사의 표본추가 및 가중치 산출

박인호* · 김기민** · 김정우***

패널조사는 특정 시점의 모집단을 대표하도록 추출한 표본 개체를 다수 시점에 걸쳐 추적하여 조사하므로 해당 시점의 모집단 전체의 평균적 특성 변화인 순변동은 물론, 개인 수준의 특성 변화인 총변동의 추정을 가능하게 한다. 하지만 계속되는 조사로 발생하는 무응답, 적격성 상실, 추적 실패 등이 누적되는 패널마모를 피할 수는 없다. 특히 사업체 대상의 패널조사는 휴·폐업과 합병, 신규 사업장 생성 등과 같은 구성 변동, 종사자나 매출액 등의 사업체 규모 변동, 그리고 업종 변경 및 지역 이전 등과 같은 특성 변동이 다소 극단적으로 발생한다.

패널자료를 이용하여 추적 시점의 횡단면 모집단 특성을 분석하는 경우에는 패널마모로 인한 대표성 보강을 위해 표본추가가 종종 고려된다. 본 연구는 한국노동연구원이 주관하는 사업체패널조사의 2016년도 제6차 조사에서 고려한 표본추가를 위해 활용된 이중틀 접근에 대해 살펴본다. 이중틀 접근은 주로 횡단면 모집단 분석을 위한 표본설계에서 서로 다른 포함율을 갖는 두 개의 표본추출틀을 활용한 표본추출기법이다. 하지만 시간흐름에 따른 모집단 구성 변동이 심한 사업체패널인 경우에 이중틀 접근을 활용하면 사업체등록정보가 없는 한계적인 상황에서 보다 의의적으로 모집단 변화를 반영하는 표본추가를 수행할 수 있다. 또한 의사최대가능도 추정을 적용하면 기존 횡단면 표본은 물론 표본추가로 확대된 횡단면 표본에 대해서도 적절한 가중치 산출이 가능하다.

주제어: 사업체패널조사, 이중틀, 표본추가, 종단면가중치, 횡단면가중치

* 부경대 통계학과 부교수(ipark@pknu.ac.kr), 교신저자.

** 한국노동연구원 전문위원(kimin1104@kli.re.kr).

*** 한국노동연구원 전문위원(kjw@kli.re.kr).

I. 서론

패널조사(panel survey)는 특정 시점의 모집단을 대표하도록 추출한 표본 개체를 다수 시점에 걸쳐 추적하여 조사하므로 해당 모집단의 시점에 따른 평균적인 특성 변화인 순변동(net change)은 물론, 개체 수준의 특성 변화인 총변동(gross changes)의 추정도 가능하다. 패널조사는 자료분석 방식의 다양성으로 인해 시점별 횡단면에 대한 반복조사(repeated cross-section survey)에 비해 선호되기도 한다.

하지만 추적조사에서 발생하는 무응답, 적격성 상실, 추적 실패에 의해 누적되는 패널마모(panel attrition)를 피할 수는 없다. 특히 사업체들을 대상으로 하는 패널조사에서는 휴·폐업과 합병은 물론, 신규 사업체 생성 등과 같은 구성 변동, 종사자나 매출액 등 사업체 규모의 확대와 축소, 그리고 업종 변경 및 지역 이전 등과 같은 특성 변동이 다소 극단적으로 발생할 수 있다. 패널조사에서는 종종 고정패널(fixed panel) 방식이 고려되는데 최초 연도의 조사자료에 대한 횡단면 분석과 이후 추적시점의 조사자료에 대한 종단면 분석이 조사연구의 주요 관심대상이 된다. 가구를 대상으로 하는 패널조사에서는 많은 경우 추적조사 시점의 횡단면 모집단에 대한 자료분석을 고려하며, 이로 인해 일정 시간이 경과한 시점에서의 모집단 변동과 누적된 패널마모로 인한 정도 수준 개선을 목적으로 표본추가를 수행한다(변종석·박인호 2019). 사업체를 대상으로 한 패널조사는 앞서 언급한 대로 가구대상의 패널조사에 비해 다소 이른 시기에 추적조사 시점의 횡단면 모집단 분석의 대표성 보강을 위한 표본추가가 고려될 수 있다.

본 연구는 한국노동연구원이 주관하는 사업체패널조사의 2016년 제6차 조사에서 수행한 표본추가에서 활용한 이중틀(dual frame) 접근에 대해 살펴본다. 사업체패널조사는 우리나라 30인 이상 규모의 사업체를 대표하는 패널조사이며 격년 주기로 수행한다. 이 조사는 우리나라의 고용구조 및 노동수요를 체계적으로 파악하고, 전반적인 인적자원 관리체제와 노사관계를 평가하여 바람직한 노동시장 및 노사관계 정책을 수립하기 위해 시행된다. 2006년 패널구축 이후 격년 주기로 2014년까지 총 5차례 조사가 진행되었으며, 4차례에 걸친 추적조사를 통해 패널이탈과 표본대체(substitution)가 누적되어 악화된 표본대표성을 개선하고 조사연구의 지속 가능성을 향상시킬 수 있는 표본추가가 요구된 바 있다.

본 논문의 구성은 다음과 같다. 제2장에서는 기존 사업체패널조사의 표본구축과 패널이탈에 대해 간단히 소개한다. 제3장에서는 사업체패널조사의 특성을 감안하여 시간흐름에 따른 모집단 구성 변동을 반영할 수 있는 이중률 접근을 통한 횡단면 표본추가에 대해 논의한다. 제4장에서는 기존의 종단면 가중치 산출은 물론 이중률 접근을 통한 횡단면 표본추가를 적절히 반영할 수 있는 횡단면가중치 산출방법에 대해서 논의한다. 횡단면가중치는 이중률 접근을 반영할 수 있는 의사최대가능도 추정(pseudo maximum likelihood estimation)을 적용하였다. 제5장은 결론 및 향후 연구주제에 대하여 간단히 논의한다.

II. 기존 사업체패널조사의 표본구축 및 패널이탈 현황

사업체패널조사(Workplace Panel Survey, WPS)¹⁾의 목적은 우리나라의 사업체를 둘러싼 내·외부 노동시장의 구조와 실태를 파악하는 데 있다. 구체적으로 고용구조와 인적자원 관리 등 내부노동시장, 그리고 노사관계의 현황이 주요 조사대상이다.

조사내용을 좀 더 상세히 살펴보면 우선, 사업체 단위 노동수요의 구조와 임시·일용직 및 사내하청 등 비정규직 현황 파악 등을 통해 고용구조에 대한 종합적 판단의 기초자료를 마련하고자 한다. 다음으로, 임금체계, 채용, 승진, 평가, 보상, 퇴직 등 인적자원 관리구조 및 특성과 노조활동, 단체교섭, 노사협의, 파업 등 노사관계의 흐름과 변화를 파악하고자 한다. 마지막으로, 교육훈련과 숙련 형성 등 인적자원 개발 및 투자의 실태 및 변화를 파악한다.

사업체패널조사는 먼저 사업체의 인력 현황 및 재무 현황과 관련한 조사를 서면으로 진행한 후, 일반적인 사업체 특징, 고용 현황 및 고용관리, 보상 및 평가, 인적자원 관리 및 작업조직, 인적자원 개발, 기업복지, 산업재해 등을 조사하고, 노무부서 현황, 노동조합 현황, 노사관계 일반 현황, 노사협의회 현황, 경영참여 등의 내용을 조사한다. 인력 현황과 재무 현황을 제외한 모든 조사는 조사대상 사업체를 방문하

1) 사업체패널조사는 2002년, 2003년, 2004년의 사전조사를 바탕으로 2006년부터 격년 주기로 시행되었는데 이를 1차년도 조사라 칭한다. 1차년도 조사인 WPS2005는 2005년 말의 정보를 기준으로 2006년에 조사된 것이고 WPS2007(2차년도 조사), WPS2009(3차년도 조사), WPS2011(4차년도 조사), WPS2013(5차년도 조사), WPS2015(6차년도 조사), WPS2017(7차년도 조사)은 각각 2008년, 2010년, 2012년, 2014년, 2016년, 2018년에 조사되었다(현재 8차년도 조사인 WPS 2019가 조사 중이다).

여 컴퓨터를 이용한 대면조사기법인 CAPI(Computer Assisted Personal Interviewing)를 활용한다.

표본추출틀로는 패널구축 당시 사업체 관련 통계청의 최신 자료인 ‘사업체기초통계조사’²⁾ 자료를 이용하였다. 조사대상으로 일반사업장 부문은 농·임업, 어업, 광업 등의 산업을 제외한 상용근로자 30인 이상 사업장, 공공부문은 20인 이상인 기관을 포함하였다. 다만, 일반적 사업장과 성격이 크게 다른 종교단체, 아파트 관리사무소, 주점업, 미용, 욕탕 및 유사 서비스업, 회원단체 등과 비법인 단체 등은 제외하였다. 산업별, 사업장 규모별, 지역별로 각각 신뢰성 높은 통계생산을 위해 관련 특성에 기초한 모집단 층화를 결정하였다. 지역 구분으로 서울권, 경기·인천권, 강원·충청권, 전라·제주권, 영남권 등의 5개 권역, 산업특성으로는 12개 업종, 사업체 규모로는 30~99인, 100~299인, 300~499인, 500인 이상의 구분을 고려하였다. 표본수와 표본배분은 모비율 추정치 정도수준을 고려하여 95% 신뢰수준에서 전국 통계는 3%, 업종별로는 8~12% 수준을 만족할 수 있도록 결정하였다. 전체 표본사업체는 총 4,275개 사업장으로 일반부문 3,916개와 공공부문 359개로 구성되었고 이 중 일부 조사실패(일반부문 41.1%, 공공부문 10.6%)와 부적격(일반부문 17.7%, 공공부문 8.6%)을 제외하고 각각 1,615개와 290개의 응답(표본) 사업장들이 최초 패널로 구축되었다. 표본추출틀 정보에 비해 조사 시점에서 사업장 규모가 축소되거나, 휴·폐업 등의 사유가 발견되었고 이러한 정도는 업종별로 다소 상이하였다(김수진·김기민 2007: <표 11>).

통계적 타당성을 갖춘 사업체단위의 분석을 가능하게 하기 위해 1차년도 횡단면가중치와 추적조사의 패널가중치를 각각 산출하였다. 먼저, 1차년도 횡단면가중치는 패널구축 당시의 조사 시점인 2005년 12월 말 기준 상용근로자 30인³⁾ 이상 전체 사업체 모집단을 대표하도록 1차년도 응답 사업체들에 대해 산출하였다. 이를 위해 복합표본설계 및 조사과정을 적절히 반영할 수 있도록 추출확률, 무응답 및 사후층화 조정은 물론 극단값(outlier) 조정 등을 포함하여 계산하였다.

패널가중치(panel weight)는 1차년도를 기준으로 상용근로자 30인 이상인 표본 응답사업체를 대상으로 하고, 이후 $t(\geq 2)$ 차년도 조사에서 해당 사업체의 상용근로자 30인 여부에 관계없이 조사 적격으로 고려하여 산출하였다. 우선 $(t-1)$ 시점의 패널표본 응답사업체는 t 시점의 추적조사에서 다음과 같이 서로 겹치지 않은 사업

2) 2004년 12월 말 기준의 사업체기초통계조사이고 현재 ‘전국사업체조사’로 명칭이 변경되었다.

3) 공공부문은 상용근로자 20인 이상을 대상으로 조사하였다.

체들로 분해할 수 있다.

$$s_{t-1}^{PR} = s_t^{PR} \bigcup^d s_t^{PNR} \bigcup^d s_t^{PIE} \quad (1)$$

여기서 s_{t-1}^{PR} 과 s_t^{PR} 는 각각 $(t-1)$ 시점과 t 시점의 패널응답이고 s_t^{PNR} 와 s_t^{PIE} 는 t 시점의 패널무응답과 패널부적격으로 $s_{t-1}^{PR} \supset s_t^{PR}$ 이고, s_1^{PR} 은 1차년도 응답사업체들을 나타내며 \bigcup^d 는 상호배반적인 두 개 집합의 합을 나타낸다. t 시점의 패널가중치는 $(t-1)$ 시점의 패널가중치에 t 시점에서의 패널응답성향점수의 역수를 곱하여 다음과 같이 산출하였다.

$$pw_{12 \dots t, i} = \frac{pw_{12 \dots (t-1), i}}{\hat{\phi}_t(x_{t-1, i})} \quad (2)$$

여기서 $pw_{12 \dots t, i}$ 는 표본사업체 i ($i \in s_{t-1}^{PR}$)의 t (≥ 2)시점 패널가중치이고, $pw_{1, i} = cw_{1, i}$, $cw_{1, i}$ 는 1차년도 횡단면가중치를 나타낸다. 또한 $\hat{\phi}_t(x_{t-1, i})$ 는 $(t-1)$ 시점 (응답)사업체 특성 $x_{t-1, i}$ 을 기반으로 추정된 t 시점의 패널응답성향점수(panel response propensity score)를 나타낸다. 사업체조사에서는 로지스틱회귀모형(logistic regression model)을 이용하였고 사업체들의 산업분류, 지역, 근로자 수, 노조유무, 설립년도, 본·지사 구분 등의 정보를 설명변수로 활용하였다.

<표 1> 1-5차 사업체패널조사의 전체 사업체에 대한 응답유형 현황 (단위: 개소)

	1차	2차	3차	4차	5차
패널표본	1,749	1,415	1,229 (69) ²⁾	1,091 (99) ²⁾	960 (85) ²⁾
[2차]		320	279	214	186
[3차]			160	128	105
[4차]				238	206
[5차]					226
특별표본 ¹⁾³⁾	156				7
합계	1,905	1,735	1,737	1,770	1,775

주: 1) 1차년도 조사에서는 공공부문에 대한 특별조사가 수행되었음.

2) 패널표본 중 괄호 안은 패널표본이 중도 이탈되었으나 그후 조사는 수행된 경우임.

3) 5차년도 추가 표본 중 7개는 1차년도의 공공부문 특별표본이었음.

사업체패널조사가 시작된 이후 휴·폐업, 응답거절 등의 패널탈락(panel attrition)은 계속적으로 진행되었다. 추적조사 시점의 종단면 자료분석과 더불어 횡단면 자료분석을 가능할 수 있게 하기 위해 다소의 패널이탈을 보완하는 표본추가, 즉 무응답 대체(substitution)를 지속적으로 실시하였다. 예를 들면, 추적조사에서 발생한 휴·폐업 및 추적불가와 응답거절로 인한 표본이탈 사업체 수의 약 4배수 예비표본을 포함하여 추출한 뒤, 기존 응답업체와의 중복업체를 먼저 제외하고 조사를 진행한 후 추가표본으로부터 산업구분, 규모, 시도 등의 표본층 구분을 만족하는 표본사업체들로 대체하였다. 예로, <표 1>은 5차년도까지의 응답유형별 표본사업체 현황과 표본추가에 의한 추가응답 수를 요약하고 있는데, 2차년도는 320개, 3차년도는 160개, 4차년도는 238개, 5차년도는 233개(민간 226개+공공 7개)의 응답사업체가 각각 추가되었음을 알 수 있다(이기재 2015).

간단히 정리하면, t 시점의 횡단면분석을 위한 표본응답 사업체는 다음과 같이 서로 겹치지 않은 사업체들로 분해할 수 있다.

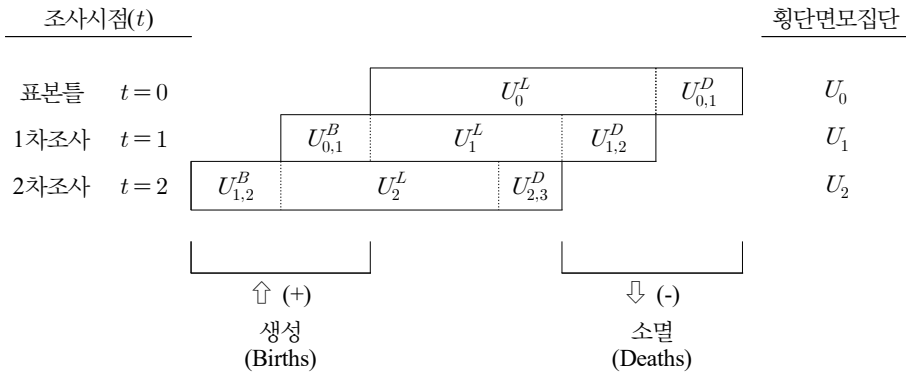
$$s_t^{CR} = s_t^R \bigcup_{d=1}^t s_d^{AR} \quad (3)$$

여기서 s_t^{CR} 는 t 시점 횡단면조사의 응답표본이고 s_t^R 는 1차년도 응답표본 s_1^R 와 $(t-1)$ 차년도 추가응답표본 s_{t-1}^{AR} 중에서, t 시점에 횡단면조사 기준에 적합한 응답사업체이며 s_t^{AR} 는 t 차년도에 추가된 응답표본을 각각 나타낸다.

1차년도와 마찬가지로 t 차년도 횡단면가중치 산정의 대상은 원칙적으로 해당 조사 기준시점에서 상용근로자 30인 이상인 사업장들이다. 하지만 조사 기준시점과 실제 조사시점 간의 시차가 있어, 상용근로자 30인 이상 사업체로 국한할 경우는 응답업체의 정보 중 상당 부분을 활용할 수 없다는 점을 고려하여 2차년도 이후 횡단면 조사에서는 상시근로자⁴⁾ 25인 이상인 표본사업체를 횡단면가중치 산출의 대상으로 삼았다(이기재 2015). $t(\geq 2)$ 차년도 조사 기준시점 모집단에 대한 횡단면가중치 산출은 1차년도 횡단면가중치 산출과 유사한 절차를 거쳤다. 다만 패널이탈 및 표본추가에 대해 추적조사 시점을 기준으로 파악한 모집단 표본층으로부터의 표본추출을 가정하고, 추적조사 기준시점의 모집단 산업분류 및 사업장 규모별 사업장과 상시근로자 수 현황을 이용하여 사후층화를 수행하였다.

4) 사업체패널조사에서는 전체근로자(정규직+기간제+단시간 근로자)와 일용근로자를 합한 근로자를 상시근로자로 정의한다.

일반적으로 모집단은 개체의 구성과 특성 등 다양한 측면에서 시간흐름에 따른 지속적인 변화가 발생한다. 특히 개인 혹은 가구를 대상으로 하는 사회조사와 비교하면, 사업체들은 시간흐름에 따른 변화 정도가 매우 심하게 나타난다. 예를 들어, 새로운 사업체가 생성(birth)되거나, 폐업·휴업·합병 등으로 소멸(death)됨에 따라 모집단의 개체 수 자체가 변화될 수 있고, 고용규모의 확대나 축소와 같은 고용변동이나 업종 변경과 같은 개체의 특성변화도 발생된다. <그림 1>과 <표 2>는 생멸에 따른 모집단의 구성변화 및 구성요소를 도식화하여 나타내고 있다. 최초년도 표본 s_1 은 조사시점($t = 1$) 이전에 구성된 표본추출틀 U_0 로부터 추출되고 표본추출틀의 생성시점 이후 생성된 개체 $U_{0,1}^B$ 는 1차년도 조사에 포함될 수 없다.



<그림 1> 생멸에 의한 모집단 변동

<표 2> 생멸에 따른 모집단 변동

시점(t)	1차년도	2차년도
횡단면	$U_1 = U_{0,1}^B \cup U_1^L \cup U_{1,2}^D$	$U_2 = U_{1,2}^B \cup U_2^L \cup U_{2,3}^D$
패널	-	$U_{12}^L = U_{0,1}^B \cup U_1^L$
모집단 구성		
생성	$U_{0,1}^B$	$U_{1,2}^B$
소멸	$U_{0,1}^D$	$U_{1,2}^D$

반면 표본추출틀에는 포함되었지만 조사시점에 더 이상 존재하지 않는 소멸개체 $U_{0,1}^D$ 는 조사를 통해 파악할 수 있다. 추적조사가 수행되는 2차년도에서는 1차년도

이후의 생성과 소멸에 의한 모집단 구성변동의 내역은 유사하게 정의된다. 2차년도 종단면 분석대상인 패널모집단은 시점별 모집단 $U_t (t = 1, 2)$ 에 모두 존재하는 개체 $U_{12}^P = U_{12}^L = U_1 \cap U_2$ 으로 정의되며⁵⁾ t 차년도는 다음과 같이 일반화하여 정의할 수 있다.

$$U_{12 \dots t}^P = \bigcap_{k=1}^t U_k \quad (4)$$

따라서 패널가중치의 작성 대상자는 식(1)의 s_t^{PR} 으로 식(4)의 패널모집단 $U_{12 \dots t}^P$ 에 대한 대표성을 갖는다. 반면, $t (\geq 2)$ 차년도 횡단면가중치의 작성 대상자는 식(3)의 s_t^{CR} 으로 횡단면 모집단 U_t 을 적절히 대표하기 위해서는 s_t^R 와 s_t^{AR} 의 추출 확률 및 응답확률에 대한 적절한 산출이 필요하다.

사업체패널(WPS)과 유사한 해외사례로는 영국의 WERS(Workplace Employment Relations Survey)와 캐나다의 WES(Workplace and Employee Survey) 등이 있다. 먼저, 영국의 WERS는 횡단면 분석과 종단면 분석을 위한 표본을 각각 구분하여 추출한다. Chaplin et al.(2005)에 의하면 t 차년도 WERS의 횡단면 표본은 조사기준의 사업체등록(business register) U_t 로부터 기존 패널표본을 제거하고 추출한다. 패널표본은 최초년도 조사설계 시점의 사업장 규모(예, U_0)에 근거하여, 당시 표본층별로 동일한 추출률에 의한 단순층화추출로 선택한다. 반면, 캐나다의 WES는, Patak et al.(1998)에 의하면, 최초년도 표본 s_1 은 표본추출률 U_1 으로부터 확률추출하고 이후 2년 간격으로 추적조사 시점 s_1 으로부터 확률적으로 s_{12}^L 을 구축한다. 더불어 새롭게 생성된 사업체 $U_{1,2}^B$ 로부터 일정규모의 표본 s_2^B 를 뽑아 생성패널에 추가하여 추적조사 시점의 횡단면 모집단 U_2 의 대표성을 확보하였다. 즉, 앞서의 해외사례는 모두 표본추출률과 표본추출을 사업체 생멸 및 시점 간 변동을 적절히 반영하도록 관리하여 패널모집단은 물론 조사시점의 횡단면 모집단을 잘 반영하도록 하고 있음을 알 수 있다.

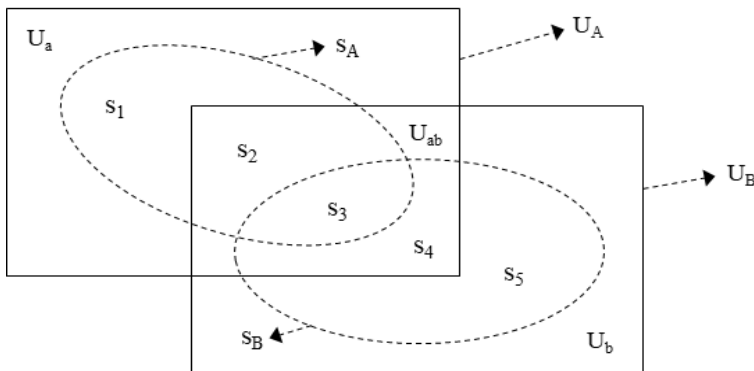
다음 장에서는 6차년도 사업체패널조사의 표본추가 방식에 대한 개선사항을 논

5) 논리적으로는 $U_{0,1}^B$ 도 패널모집단으로 포함시킬 수 있지만 조사에서 파악은 불가능하고 사후층화를 통해서만 반영이 가능하다. 예외적으로 사업체등록부와 같이 시점별 사업체 명단을 파악하는 명부가 존재한다면 이후 필요 시 해당 부분을 표본에 포함시킬 수도 있을 것이다.

하고 현재 한국 통계 인프라를 고려하되 앞서 논의한 현행의 문제점을 보완할 수 있는 접근에 대해 논의하고자 한다.

Ⅲ. 이중틀 적용을 통한 표본추가

일신된 사업체패널조사 표본을 기반으로 한 새로운 웨이브는 2006년, 2008년, 2010년, 2012년, 2014년까지 다섯 번 조사되었다. 조사가 진행되면서 표본이탈과 이에 따른 표본대체(substitution)의 효과가 누적되어 표본대표성 확보에 대한 필요성이 대두되었다. 이러한 목적을 달성하기 위해 추가될 표본사업체는 기존 패널모집단의 특성은 물론 가급적 표본추가 시점의 횡단면 모집단 특성 또한 반영할 수 있는 방법을 모색하여야 한다. 이를 위해 <그림 2>와 같이 이중틀(dual frame) 방법을 고려하였다. 2005년과 2015년의 횡단면 모집단을 각각 U_A 와 U_B 로 표기하고, 추적기간 동안 조사적격성을 유지한 사업체와 소멸한 사업체를 각각 U_{ab} 와 U_a 로, 2005년 이후 생성된 사업체는 U_b 로 표기하자.



<그림 2> 목표모집단 및 표본 분해도

2005년 횡단면 모집단은 $U_A = U_a \cup U_{ab}$ 로, 또한 2015년 횡단면 모집단은 $U_B = U_b \cup U_{ab}$ 로 각각 표기될 수 있다. 더불어 2005년 횡단면 모집단은 휴·폐업에 따른 소멸분 U_a 와 적격성 유지분 U_{ab} 로 나뉘며, 후자의 경우에는 구성 및 특성 변화를

파악할 수 있는 패널조사의 대상이 된다. 2005년과 2015년의 횡단면 모집단 U_A 와 U_B 로부터 추출된 표본은 각각 s_A 와 s_B 가 되며, 두 표본에 중복된 업체 s_3 , 휴·폐업에 의한 소멸분 s_1 , 신생업체 s_5 , 중복되지 않고 양 시점에 존재하는 표본 s_2 와 s_4 로 구분할 수 있다. 두 시점에 걸쳐 존재하는 패널모집단 U_{ab} 에 속하는 표본은 s_{ab} 로 이는 s_2, s_3, s_4 로 구성된다.

이중틀 방식은 일반적으로 횡단면 조사를 위한 표본설계에서 모집단에 대해 서로 다른 포함률을 갖는 표본추출틀을 함께 고려할 때 사용한다. 하지만 본 연구에서는 사업체패널을 위한 표본추가에 이중틀 방식을 적용하였다. 이는 생성, 소멸, 분리, 병합, 규모 및 분류 변경 등과 같은 모집단 변화가 반영된 사업체명부가 관리되지 않고, 시점별 명부만 존재하는 우리나라 상황에서 제한적이나마 모집단 변동을 반영하고자 하는 노력의 일환이었다. 따라서 1~5차년도 조사를 위해 추출되었지만 무응답과 소멸이 아닌 분리, 병합, 규모 및 분류변경 등의 이유로 패널에서 이탈되었던 개체들은 추가표본에 중복적으로 포함될 수 있는 가능성이 존재함에 유의해야 한다.

표본추출틀로는 통계청의 「전국사업체조사」⁶⁾에서 파악된 사업체 명부로 1차년도 횡단면 모집단 U_A 에 대해서 2004년 12월 말 기준의 자료를 사용하였고, 6차년도 횡단면 모집단 U_B 에 대해서는 2014년 12월 말 기준의 자료를 사용하였다. 표본추가에 고려한 표본추출틀의 구성 기준시점은 6차년도 사업체패널조사의 조사시점과는 약 1년 이상의 시차가 발생하므로, 1차년도 패널구축에서와 마찬가지로 표본추출틀이 구성된 이후 추가로 생성된 신규사업체가 표본추출에서 누락되는 한계는 여전히 존재한다. 6차년도 사업체패널조사의 횡단면 모집단에 대한 표본추출틀은 2014년 12월 말 기준의 「전국사업체조사」의 사업체 명부 중(금융·보험업, 전기·가스·수도업 등의) 일부 업종을 제외한 사업체들이며, 이중 조사대상 사업체는 상시 근로자 30인 이상인 사업체 중 조사수행, 자료분석 등을 고려하여 일부 분류를 추가로 제외한 총 57,894개⁷⁾ 사업체이다. 단, 고용보험 DB와의 연계에 대해서는 고려하지 않았다(박인호·김기민 2016).

6) 「전국사업체조사」는 1994년에 「전국 사업체기초통계조사」라는 명칭으로 처음 실시되었고, 2008년에 현재의 명칭으로 변경되어 실시되고 있는 조사이다. 매년 조사기준일(12월31일)을 기준으로 우리나라에서 산업활동을 수행한(또는 수행하고 있는) 종사자 1인 이상 모든 전국 사업체를 대상으로 하는 조사이다. 이 조사는 전국의 지역별 사업체의 규모 및 분포를 파악하여 정부의 정책수립 및 평가, 기업의 경영계획 수립 및 학술 연구 등의 기초자료로 활용되며, 다른 사업체단위 통계조사의 표본추출틀로 활용되고 있다.

7) 이중 사업체패널조사에 포함되어 있는 기존 사업체 수는 총 1,451개소이다.

모집단 층화는 산업분류와 사업체 규모를 각각 10개와 4개로 구분하여 최대 40개 층으로 구성하였다. 산업분류는 제조업, 건설업, 전기·가스·수도업의 경우 과거 사업체패널조사의 분류와 동일하게 정하였고, 서비스업의 경우에는 OECD 분류기준을 적용하여 4가지로 분류⁸⁾하였다. 사업체 규모는 상시근로자, 즉 상용근로자와 임시·일용근로자 기준으로 30~99명, 100~299명, 300~499명, 500인 이상의 4가지 분류를 고려하였다.

6차년도 사업체패널조사의 추가표본을 위한 표본추출에서는 기존 표본설계에서 고려하였던 지역배분은 제외하고 산업별 배분과 규모별 배분을 순차적으로 고려하되, 기존 경험을 바탕으로 예측응답률을 고려하여 최종적인 표본배분을 수행하였는데 요약하면 다음과 같다(박인호·김기민 2016).

- i) 업종별로 배분하되 전기·가스·수도업은 별도 추출(25개 고정⁹⁾)
- ii) 업종 내 500인 이상 규모층에 대해 전수조사(약 39% 응답률¹⁰⁾ 가정)
- iii) 500인 미만 규모층에 대해 표본조사 적용
- iv) 표본 제공방식은 예측응답률¹¹⁾을 고려하여 결정

표본배분은 산업분류와 사업체규모를 순차적으로 고려하였고 기본적으로 비례배분과 균등배분의 절충형태인 키쉬배분(Kish allocation)을 적용하였다. 신규표본은 실질적으로는 두 차례에 걸쳐 추출되었는데, 이는 조사환경 악화에 따라 기존 조사의 응답률을 적용한 앞서의 표본배분으로는 목표응답 수를 채울 수 없어서 2차 신규표본이 추가되었기 때문이다. 2차 신규표본은 Keyfitz 방식(Brick et al. 1987)의 조건부 추출확률이 적용되었고 결과적으로 무조건부 추출확률이 층별로 동일할 수 있도록 하였다. <표 3>은 사업체 규모 및 산업분류에 따른 전체 신규표본의 규모를

-
- 8) OECD에서 정의하고 있는 4가지 범주는 사업서비스업(producer service), 유통(배분)서비스업(distributive service), 개인서비스업(personal service)과 사회서비스업(social service)로 구성된다. 사업서비스업은 금융 및 보험업, 부동산 및 임대업, 사업서비스업을 포함하고 유통(배분)서비스업은 운수업, 통신업을 포함하며, 개인서비스업은 도소매, 숙박 및 음식점, 오락문화 및 운동관련업, 가사 서비스업, 기타 공공수리 및 개인서비스업이며, 사회서비스업은 공공행정, 국방 및 사회보장행정, 보건 및 사회복지사업, 교육 서비스업, 국제 및 외국기관으로 구성된다(8차 한 국표준산업분류에 의거).
 - 9) 과거 25개 기준에 대한 이유를 찾지 못하여 해당 업종의 주요한 추정량들의 정도수준(RSE 등)을 산출하여 판단하였다.
 - 10) 2006년에 실시한 사업체패널조사에서 500인 이상 규모의 사업체 평균 응답률이 약 39%로 나타났다. 여기에 0.8를 곱하여 보수적 수치를 적용하였다.
 - 11) 예측응답률은 2006년 조사경험을 토대로 추정하였다. 여기에 0.8를 곱하여 보수적 수치를 적용하였다.

보여 준다.¹²⁾

<표 3> 6차년도 신규표본 사업체 현황(1,2차 통합)

(단위: 개소)

산업분류	30~99인	100~299인	300~499인	500인 이상	계
경공업	492	367	59	29	947
화학공업	494	251	67	63	875
금속·자동차·운송	1,271	676	166	118	2,231
전기·전자·정밀공업	688	500	89	106	1,383
건설업	1,248	803	115	92	2,258
개인서비스업	1,377	428	120	70	1,995
유통서비스업	492	256	72	47	867
사업서비스업	1,109	864	510	507	2,990
사회서비스업	446	265	71	117	899
전기·가스·수도업	25	26	5	2	58
소계	7,642	4,436	1,274	1,151	14,503
공공부문	6	10	19	13	48
계	7,648	4,446	1,293	1,164	14,551

<표 3>에는 신규표본의 표본층인 사업체 규모 및 산업분류에 따른 표본사업체 현황이 정리되어 있다. 공공분문은 총 48개이고 비공공분문은 총 14,503개 사업체로 이루어졌다. <표 4>는 신규표본의 사업체 규모 및 산업분류에 따른 응답사업체 현황을 정리하고 있다.

<표 5>는 6차년도 조사에서 표본구분 및 응답유형별 표본사업체의 현황을 정리하고 있다. 기존 패널표본은 총 1,669개 사업장에 포함되었는데, 이 중 6차년도 조사기준으로 소멸된 사업장 s_1 은 50개소, 신규표본과 중복되지 않는 사업장 s_2 는 3건의 중복보고를 포함하여 총 899개소, 신규표본과 중복된 사업장 s_3 는 총 720개소이다. 신규표본은 총 13,831개소로 이 중 패널표본과 중복되지 않은 사업장 중 2013년 이전에 설립된 사업장 s_4 는 총 13,304개소¹³⁾이고 2014년에 설립된 사업장 s_5 는 총 527개소이다.

12) 1차, 2차 구분의 표본응답 수는 생략하였고, 과거 표본과 중복되는 사업체는 포함하였다.

13) 패널표본 중 조사에 참여하고 있지 않은 표본사업장일 수는 있으나 6차년도 조사에서 확인이 불가능한 사업장을 포함할 수도 있다.

<표 4> 6차년도 신규표본 횡단면조사 산출 대상 응답사업체 현황 (단위: 개소)

산업분류	30~99인	100~299인	300~499인	500인 이상	계
경공업	141	88	31	7	267
화학공업	114	48	33	24	219
금속·자동차·운송	200	78	68	35	381
전기·전자·정밀공업	110	60	41	14	225
건설업	116	71	28	17	232
개인서비스업	189	88	41	16	334
유통서비스업	98	73	41	6	218
사업서비스업	172	149	96	79	496
사회서비스업	120	48	35	30	233
전기·가스수도	11	12	3	-	26
소계	1,271	715	417	228	2,631
공공부문	3	10	15	11	39
계	1,274	725	432	239	2,670

주: 횡단면조사 산출 대상 응답사업체는 상시근로자 25인 이상인 경우임.

<표 5> 6차년도 표본종류별 응답현황 (단위: 개소, %)

표본구분	적합				부적합		합계 (T)
	응답 (A)	무응답 (B)	소계 (C=A+B)	응답률 (A/C)	표본사업체 (D)	부적합비율 (D/T)	
기존	s_1	-	-	-	50	100.0%	50
	s_2	709	139	848	83.6%	51	899
	s_3	534	166	700	76.3%	20	720
	소계	1,243	305	1,548	80.3%	121	1,669
추가	s_4	2,076	9,911	11,987	17.3%	1,317	13,304
	s_5	60	390	450	13.3%	77	527
	소계	2,136	10,301	12,437	17.2%	1,394	13,831
합계	3,379	10,606	13,985	24.2%	1,515	9.8%	15,500

주: 상시근로자 수가 25인 이상인 민간 및 공공부문 사업체에 대한 정보임.

IV. 가중치 산출

사업체패널조사를 통해 조사된 자료 분석에는 (i) 기존 1차년도 이래 지속된 패널표본에 대한 종단면 분석과 (ii) 기존 표본과 추가된 표본을 혼합한 확대표본에 대한 횡단면 분석, 그리고 (iii) 향후 확대표본에 대한 종단면 분석 등이 포함될 수 있다. 본 연구에서는 사업체패널조사에서 산출하는 두 가지 표본가중치인 (i) 종단면 분석을 위한 패널가중치와 (ii) 횡단면 분석을 위한 횡단면가중치의 산출에 대해 간략히 소개한다.

1. 패널가중치

사업체패널조사의 종단면 모집단은 1차년도 사업체패널조사의 조사기준 시점인 2005년도 12월 말 기준으로 상용근로자 30인 이상의 우리나라 전체 사업체이다. 종단면 분석의 목적은 추적조사를 통해 해당 사업체들의 전반적인 경영환경 및 인적 자원 관리체계, 노사관계의 현황 및 임금교섭 과정 등이 어떻게 변화하고 있는지를 파악하고자 함이다. 따라서 패널가중치는 기존 방식인 식(2)와 동일하게 산출하였다.

2. 표본추가를 반영한 횡단면가중치

횡단면 표본사업체는 이중틀 접근을 통한 표본추가로 인해 기존표본 s_A 와 신규표본 s_B 로 구성되었고 이에 따라 표본별 횡단면가중치가 각각 먼저 구축되었다. 기존표본과 신규표본의 횡단면가중치 산정은 기존방식(이기재 2015)과 동일하게 적용하였다. 최초년 혹은 2015년 표본추출시 사용하였던 산업대분류와 사업체 규모를 표본층으로 하는 복합표본설계(complex sample design)를 반영한 ㉠ 설계가중치, ㉡ 무응답 조정, ㉢ 사후층화 조정, ㉣ 극단값 조정 등의 네 가지 구성요소를 통합하여 산정하였다. 기존표본과 신규표본의 횡단면가중치는 의사최대가능도(擬似最大可能度, pseudo-maximum likelihood, 이후 PML) 추정(Skinner & Rao 1996)에 근거하여 산출하였다. 따라서 6차년도 조사의 횡단면가중치는 다음과 같은 세 종류로

이루어진다.

- ① 기존표본 $s_A (= s_1 \cup s_2 \cup s_3)$ 에 대한 횡단면가중치 w_{Ak}
- ② 신규표본 $s_B (= s_3 \cup s_4 \cup s_5)$ 에 대한 횡단면가중치 w_{Bk}
- ③ 전체표본 $s (= s_A \cup s_B)$ 에 대한 PML 횡단면가중치 $w_{PML,k}$

6차년도 사업체패널조사를 위해 고려된 기존표본은 총 1,669개 사업체(<그림 2>, <표 5> 참고)이며 이 중 s_2 의 3개 사업체는 중복 보고되었으므로 이를 제거하면 총 1,666개가 기존표본 횡단면가중치의 산출 대상이다. 기존표본 s_A 는 기본적으로 기존모집단 U_A 로부터 추출된 사업체이다. 하지만 3장에서 기술된 것처럼 1차년도 조사를 위한 최초표본은 물론 2~5차년도 조사 당시 기준으로 신생 사업체들을 일부 포함하고 있다(<표 1> 참고). 이중틀 접근의 일반적 형태와는 달리 기존 조사에 대한 모집단은 2005년 이후 2015년까지의 생멸을 포함하는 사업체들로 구성된다. 이는 패널이탈로 인한 지속적인 표본사업체의 감소를 방지하고 사업체패널조사의 횡단면적 추정이 가능하도록 하기 위하여 2차년도 조사부터 표본사업체를 추가하여 사업체패널을 보완해 왔기 때문이다(이기재 2015).

사업체패널조사에서 횡단면가중치 산정의 대상은 원칙적으로 해당 조사 기준시점에서 상용근로자 30인 이상인 응답사업체이다. 또한 이전 차수에서 무응답한 사업체일지라도 관심 차수에 다시 참여한다면 응답업체로 분류한다. 하지만 조사 기준시점과 실제 조사시점 간 시차발생은 물론, 상용근로자 30인 이상 사업체로 국한할 경우에 응답사업체의 정보 중 상당 부분을 활용할 수 없다는 점을 고려하여 상시근로자 25인 이상인 표본사업체를 횡단면가중치 산출의 대상으로 삼았다. 2~5차년도 조사에서도 동일한 기준에 따라 횡단면가중치 산출 대상을 국한하였다. 기존 표본의 횡단면가중치는 기존 방식에 따라 설계가중치, 무응답 조정, 사후층화 조정이 순차적으로 적용되었다(박인호 2018).

6차년도 조사를 위한 신규표본은 총 14,551개 사업체(<그림 2>, <표 5> 참조)로 구성된다. 이중틀 추출방식에 따라 기존표본과 중복되어 추출된 표본 s_3 은 720개의 사업체로 이루어지며, s_4 는 기존표본 조사시점에 존재하나 기존표본에 포함되지 않고 새롭게 추출된 13,304개 사업체로 이루어지고, s_5 는 2016년 조사시점을 기준으로 한 신생업체 527개 사업체로 구성된다. 따라서 6차년도 조사에서 순수하게 추가

된 신규표본은 s_4 와 s_5 인 13,831개 사업체이다. 6차년도 조사의 횡단면가중치 산정의 대상은 기존표본과 동일하게 원칙적으로 조사 기준시점에서 상시근로자 25인 이상인 표본사업체를 횡단면가중치 산출의 대상으로 삼았다. 신규표본의 횡단면가중치는 기존표본과 동일하게 설계가중치, 무응답 조정, 사후층화 조정이 순차적으로 적용되었다. <표 6>은 사후층에 대한 정보를 나타내고 있다(박인호 2018).

<표 6> 사업체패널조사의 사후층 크기

(단위: 개소)

산업분류	사업체규모 (상시근로자수)					계
	25~49인	50~99인	100~299인	300~499인	500인 이상 ¹⁾	총계
경공업	2,412	1,375	659	60	27	4,533
화학공업	1,849	1,137	617	70	(38, 21)	3,732
금속·자동차·운송	4,442	2,905	1,468	163	(69, 49)	9,096
전기·전자·정밀공업	1,625	1,109	690	89	(53, 43)	3,609
건설업	3,018	1,864	1,048	163	145	6,238
개인서비스업	6,018	3,421	1,453	140	81	11,113
유통서비스업	1,729	1,537	1,253	90	62	4,671
사업서비스업	5,568	4,071	2,721	597	(368, 160)	13,485
사회서비스업	2,302	2,075	939	72	(56, 62)	5,506
전기·가스·수도업	27	23	27	6	3	86
계	28,990	19,517	10,875	1,450	1,237	62,087

주: 1) 500인 이상의 열에서 (,)는 (500~999인, 1,000인 이상)을 의미함.

자료: 통계청, 「전국사업체조사(2015년)」.

만약 사업체 DB가 시점별로 연계되어 생성과 소멸, 업종 및 규모 변화 등의 정보가 누적되어 사업체 고유번호로 갱신되어 관리된다면 사업체 변동을 더 잘 반영할 수 있을 것이다. 예를 들어, 미국 노동통계국(BLS, U.S. Bureau of Labor Statistics)은 표본들의 재정비는 물론 주요 조사들의 재설계를 통해 조사들 간의 표본할당 및 추출을 중앙집중식으로 처리하며, 신규업체에 대해서는 독립적으로 영구난수(permanent random number)를 발생시키고, 표본층별로 조정하여 기존표본에 추가하고 또한 기존표본 중에 폐업한 사업체는 제거하는 등 사업체 모집단의 변동을 반영하고 있다(Butani et al. 1998). 하지만 우리나라에서는 현재 사업체DB가 마련되

어 있지 않아 시점 간 사업체 변동에 대한 내역을 파악할 수 없다.

이러한 상황에서 이중틀 방식을 응용한 표본추출은 기존표본에 최신 사업체 명부를 이용한 신규표본을 추출하여 추가함으로써 제한적이거나 시점별 모집단 변동을 파악할 수 있게 하는 장점이 있다. 이중틀 방식은 두 개의 서로 다른 추출틀로부터 독립적인 표본을 추출하므로 공통부분으로부터 추출된 표본에 대해서는 각각의 추출틀로부터 포함될 수 있는 이중적인 추출확률이 감안되어 가중치가 절충적으로 산출되어야 한다. 본 연구에서는 Skinner & Rao(1996)의 의사최대가능도 가중치 방식을 적용하였다. 의사최대가능도 방식은 추출틀 크기인 N_A 와 N_B 는 물론 공통영역인 U_{ab} 의 크기추정량 \hat{N}_{ab} 등의 보조 정보를 활용하되, 특정 조사변수 y 와는 무관하게 가중치를 산출한다.

기존표본은 개념적으로 2005년부터 2015년까지의 횡단면 모집단으로부터 추출되었다고 가정할 수 있다. 하지만 실질적으로는 2005년부터 2015년까지 적격성을 유지하는 사업체 U_{ab} 는 물론 이미 폐업한 사업체 U_a 도 포함한다. 따라서 응답업체만으로는 기존패널 중 폐업한 사업체표본 s_1 을 제외한 표본 $s_{ab}' = s_2 \cup s_3$ 만을 파악할 수 있고 해당 표본 수는 $n_{ab}' = n_2 + n_3$ 이다. 표본 A 의 최종가중치 cw_{Ai}^{PS} 를 이용한 모집단 총수 추정량은 $\hat{N}_{ab}' = \sum_{i \in s_{ab}'} cw_{Ai}^{PS}$ 으로 계산할 수 있다. 반면 신규표본은 2005년부터 2015년까지 적격성을 유지하는 사업체 U_{ab} 는 물론 2015년 이후 생성된 신규사업체 U_b 로부터 추출된다. 따라서 신규표본은 $s_B = s_{ab}'' \cup s_b$ 로 2005년부터 2015년까지 적격성을 유지한 크기 $n_{ab}'' = n_3 + n_4$ 의 표본 $s_{ab}'' = s_3 \cup s_4$ 와 2015년 이후 생성된 크기 $n_b = n_5$ 의 신생업체 표본 $s_b = s_5$ 로 분리된다. 신규표본의 전체 크기는 $n_B = n_{ab}'' + n_b$ 이다. 표본 B 의 최종가중치 cw_{Bi}^{PS} 를 이용한 모집단 U_B , U_{ab} , U_b 의 크기는 각각 $\hat{N}_B = \hat{N}_{ab}'' + \hat{N}_b$ 과 $\hat{N}_{ab}'' = \sum_{i \in s_{ab}''} cw_{Bi}^{PS}$, 그리고 $\hat{N}_b = \sum_{i \in s_b} cw_{Bi}^{PS}$ 로 계산할 수 있다. 따라서 동일한 모집단 U_{ab} 에 대해 두 가지 표본 s_{ab}' 과 표본 s_{ab}'' 로부터 추정되는 모집단 크기 추정량과 표본수를 이용하여 다음과 같이 절충된 PML 횡단면가중치를 산출할 수 있다.

$$\widehat{cw}_i' = \left(\frac{\tilde{N}_{ab, PML}}{\hat{N}_{ab}'} \right) \left(\frac{n_{ab}'}{n_{ab}' + n_{ab}''} \right) cw_{Ai}^{PS}, \quad i \in s_{ab}'$$

$$\widehat{cw}_i'' = \begin{cases} \left(\frac{\widehat{N}_B - \widehat{N}_{ab,PML}}{\widehat{N}_b} \right) cw_{Bi}^{PS} & i \in s_b \\ \left(\frac{\widehat{N}_{ab,PML}}{\widehat{N}_{ab}''} \right) \left(\frac{n_{ab}''}{n_{ab}' + n_{ab}''} \right) cw_{Bi}^{PS} & i \in s_{ab}'' \end{cases}$$

여기서 $\widehat{N}_{ab,PML} = (2p)^{-1} [q - (q^2 - 4pr)^{0.5}]$ 는 U_{ab} 규모에 대한 의사최대가능도 추정량이며, $r = n_{ab}' \widehat{N}_{ab}' \widehat{N}_B + n_B \widehat{N}_{ab}'' \widehat{N}_{ab}''$, $p = n_{ab}' + n_B$, $q = n_{ab}' \widehat{N}_B + n_B \widehat{N}_{ab}' + n_{ab}' \widehat{N}_{ab}' + n_B \widehat{N}_{ab}''$ 이고, $\widehat{N}_{ab}'' = n_B \widehat{N}_{ab}'' / \widehat{N}_B$ 이다.

의사최대가능도 가중치는 기존표본과 신규표본의 표본가중치를 절충한 다음의 형태로 정의된다.

$$cw_i^{PML} = \begin{cases} \widehat{cw}_i' & , i \in s_2 \\ \lambda \widehat{cw}_i' + (1 - \lambda) \widehat{cw}_i'' & , i \in s_3 \\ \widehat{cw}_i'' & , i \in s_4 \cup s_5 \end{cases}$$

여기서 $\lambda = cv_w''^2 / (cv_w'^2 + cv_w''^2)$ 는 표본 간 절충계수에 해당하며, cv_w' 는 산업별 기존표본 가중치 \widehat{cw}_i' 와의 상대표준오차, cv_w'' 는 산업별 신규표본 가중치 \widehat{cw}_i'' 의 상대표준오차를 각각 나타낸다. 앞서 기술한 의사최대가능도 가중치 산출방안은 각각의 표본으로부터 산출한 두 개의 공통영역의 크기추정량 \widehat{N}_{ab}' 와 \widehat{N}_{ab}'' 을 이용하여 계산한 의사최대가능도 추정량 $\widehat{N}_{ab,PML}$ 의 점근적 분산(asymptotic variance)을 최소화하도록 한다(Lohr 2009).

이중틀(dual frame) 의사최대가능도 가중치의 산출은 안정성 확보를 위해 6차년도 신규표본 추출을 위한 표본설계 시 고려한 산업분류를 기준으로 하며, 당시 표본설계 기준의 산업분류가 없는 91개 사업체는 6차년도 조사기준 산업분류로 대체하였다. <표 7>은 이중틀 분해에 따른 표본종류와 산업분류별 응답사업체의 현황을 나타낸다. 휴·폐업에 따른 소멸부분인 s_1 은 정의상 응답사업체는 존재하지 않으며 응답 신생업체 s_5 는 60개소이다. 생멸부분이 아닌 기존표본과 신규표본에 중복되는 응답표본 s_3 은 495개 업체이고, 기존표본과 신규표본에만 속하는 응답표본 s_2 와 s_4 는 각각 649개와 2,076개소들로 구성되었다. <표 8>은 산업분류별 PML 횡단면가중치 산출에 필요한 구성요소들을 정리하여 주고 있다.

<표 7> 이중틀 표본종류와 산업분류별 응답사업체 현황

(단위: 개소)

산업분류	표본종류					합계
	S1	S2	S3	S4	S5	
경공업	-	84	55	209	3	351
화학공업	-	66	41	176	2	285
금속·자동차·운송	-	93	79	293	9	474
전기·전자·정밀공업	-	44	47	173	5	269
건설업	-	29	42	186	4	261
개인서비스업	-	103	45	273	16	437
유통서비스업	-	88	47	170	1	306
사업서비스업	-	100	91	398	7	596
사회서비스업	-	35	37	184	12	268
전기·가스·수도업	-	7	11	14	1	33
합계	-	649	495	2,076	60	3,280

주: 민간부문의 횡단면가중치를 부여하기 위해 공공부문을 제외한 상시근로자 25인 이상 사업체에 대한 것임. 공공부문이 제외되었으므로 <표 5>의 s_2 , s_3 의 응답 수보다 적게 나타남.

<표 8> PML 횡단면가중치 산출 구성요소

(단위: 개소)

산업분류	\hat{N}_{ab}'	\hat{N}_{ab}''	\hat{N}_b	\hat{N}_B	n_{ab}'	n_{ab}''	n_B	$\tilde{N}_{ab,PML}$	γ_{ab}'	γ_{ab}''	γ_b	λ
경공업	3,916.0	4,415.0	85.9	4,497.8	139	264	267	3,911.6	0.3742	0.6258	7.0188	0.3739
화학공업	3,897.2	3,580.5	30.0	3,600.6	107	217	219	3,580.5	0.3110	0.6890	0.6718	0.4212
금속·자동차·운송	9,042.4	8,898.6	211.6	9,044.3	172	372	381	8,898.6	0.3111	0.6889	0.6890	0.4089
전기·전자·정밀공업	3,288.6	3,358.1	74.7	3,432.8	91	220	225	3,288.6	0.2969	0.7031	1.9306	0.4497
건설업	6,514.7	6,291.5	119.0	6,410.6	71	228	232	6,319.4	0.2314	0.7686	0.7657	0.4401
개인서비스업	10,288.3	10,448.5	730.4	11,178.9	148	318	334	10,288.3	0.3250	0.6750	1.2193	0.4359
유통서비스업	4,632.36	4,596.0	34.7	4,630.6	135	217	218	4,609.2	0.3824	0.6177	0.6176	0.5263
사업서비스업	12,882.0	13,422.6	276.7	13,699.3	191	489	496	12,820.0	0.2905	0.7095	2.9541	0.3898
사회서비스업	5,661.6	5,164.3	341.7	5,506.0	72	221	233	5,245.0	0.2311	0.7689	0.7639	0.1286
전기·가스·수도업	58.4	79.1	3.9	83.0	18	25	26	58.4	0.4961	0.5039	6.3874	0.5000

이때 cw_i^{PML} 은 PML 횡단면가중치, cw_i^F 은 레이킹-비 조정이 된 최종 횡단면가중치, τ 는 다음의 칼리브레이션 조건을 만족시키는 해벡터(solution vector)가 된다.

$$\sum_{i \in s_{XR}} cw_i^F x_i = \sum_{i \in U_X} x_i$$

여기서 s_{XR} 과 U_X 는 횡단면 응답표본과 모집단을 각각 나타낸다.

<표 9>는 5차년도 조사의 횡단면가중치를 포함하여 기존표본 및 신규표본, 그리고 통합표본의 횡단면가중치를 비교하고 있다. 6차년도 조사를 위한 표본추가로 인해 5차 횡단면가중치와 6차 기존표본 및 신규표본의 가중치에 비해 통합표본의 PML 횡단면가중치와 레이킹-비 가중치의 변동이 많이 작아진 것을 확인할 수 있다.

<표 9> 6차년도 사업체패널조사의 횡단면가중치 분포 (단위: 개소)

구분	응답 수	평균	표준편차	변동계수	최소	25%	50%	75%	최대
5차 표본	1,484	30.6	34.2	111.9	0.4	6.1	20.0	39.2	233.5
6차 기존표본	1,144 ¹⁾	52.6	65.4	124.4	0.6	8.1	31.6	70.4	499.9
6차 신규표본	2,136	41.0	41.0	144.2	1.0	3.9	11.4	35.5	388.2
6차(PML) 통합표본	3,280	18.0	17.0	94.8	0.1	4.8	15.3	27.7	194.8
6차(레이킹비조정)	3,280	18.9	17.6	93.5	0.2	6.4	15.0	26.0	268.3
전체(공공포함)	3,380 ²⁾	18.5	17.6	95.4	0.2	5.3	14.6	26.0	268.3

주: 1) 기존표본 중 공공부문을 제외한 사업체 수입

2) 사업체패널조사는 민간부문의 경우 상시근로자 25인 이상에 대해, 공공부문의 경우 전체 기관에 대해 횡단면가중치를 부여함. 공공부문 중 상시근로자 25인 미만 기관이 1개 존재함에 따라 <표 5>의 25인 이상 응답사업체 수와 차이가 나타남.

V. 논의

사업체는 신생, 휴·폐업 및 합병 등과 같은 개체들의 구성 변화는 물론 고용 및 매출 그리고 종사자 규모 등의 특성 변동도 매우 심하게 발생할 수 있다. 따라서 이들을 대상으로 한 패널조사는 조사거부, 적격성 상실, 추적실패, 휴·폐업 등의 이유로 무시하지 못할 수준의 패널탈락의 가능성이 존재하여 패널구축 후 상대적으로

이른 시기부터 표본추가에 대한 요구가 있을 수 있다. 만약 미국, 영국 등과 같은 주요국들이 활용하는 사업체등록정보를 이용한다면 조사설계 단계에서 사업체 생멸 및 개체 특성 변화를 사전적으로 반영하는 보다 안정적인 조사 운영이 가능할 것이다. 하지만 우리나라에는 이용가능한 사업체등록명부가 제공되고 있지 않아 전통적 방식에 따른 고정시점의 표본추출틀로부터 추출된 표본사업체로 패널을 구축하고 추적하는 방식을 고려할 수밖에 없다. 본 연구에서는 이러한 한계적 상황하에서 시간 변동에 따른 모집단 변동을 반영한 표본추가의 문제를 이중틀 접근을 통하여 해결하고자 하였다. 이러한 노력에도 불구하고 앞서 논의한 바와 같이 개체들의 구성 변화와 특성 변화가 상대적으로 심한 사업체를 대상으로 한 패널조사가 갖는 한계는 피할 수 없음을 주의해야 할 필요가 있다. 통계청은 현재 전국사업체조사와 더불어 국세청의 사업자등록자료와 법인세, 사업소득신고서 등의 다양한 행정정보를 이용한 통계기업등록부(Statistical Business Register: SBR)를 구축 중이며 2021년부터는 이를 표본추출틀로 사용하여 경제부분 등록센서스를 수행할 계획인 것으로 알려져 있다. 향후 통계청이 통계기업등록부를 외부 기관에 제공한다면 이를 이용한 표본추가를 고려할 수 있을 것으로 기대한다.

참고문헌

- 김수진·김기민. 2007. “2006년 사업체패널조사의 조사 설계와 실사 결과.” 《노동리뷰》 26: 57-71.
- 박인호·김기민. 2016. “2016 사업체패널조사의 추가 표본설계 및 표본추출.” 《패널자료 품질개선 연구(VI)》. 한국노동연구원. 3-44.
- 변종석·박인호. 2019. “「한국미디어패널조사」 품질개선 컨설팅 최종결과보고서.” 한국조사연구학회.
- 이기재. 2015. “제5차 사업체패널가중치 연구 최종보고서.” 한국조사연구학회.
- 박인호. 2018. “제6차 사업체패널가중치 연구 최종보고서.” 한국조사연구학회.

Brick, J.M., D.R. Morganstein, and C.L. Wolter. 1987. “Additional Uses for Keyfitz Selection.” Proceedings of the Section on Survey Research Methods, *American Statistical Association* 787-791.

- Butani, S., K.W. Robertson, and K. Mueller. 1998. "Assigning Permanent Random Numbers to the Bureau of Labor Statistics Longitudinal(Universe) Data Base." Proceedings of the Section on Survey Research Methods, *American Statistical Association* 451-456.
- Chaplin, J., J. Mangla, S. Purdon, and C. Airey. 2005. "The Workplace Employment Relations Survey(WERS) 2004: Technical Report(Cross-section and Panel Survey)." London: National Centre for Social(NatCen) Research.
http://doc.ukdataservice.ac.uk/doc/5294/mrdoc/pdf/5294_vol1_techreport.pdf
- Lohr, S.L.. 2009. "Multiple-Frame Surveys." in Handbook of Statistics 29A, *Sample Surveys: Design, Methods and Applications*(Pfeffermann, E. and C.R. Rao eds.). Elsevier.
- Patak, Z., M. Hidioglou, and P. Lavallee. 1998. "The Methodology of the Workplace and Employee Survey." Proceedings for Survey Research Method. *The American Statistical Association* 83-91.
- Skinner, C.J. and J.N.K. Rao. 1996. "Estimation in Dual Frame Surveys with Complex Designs." *Journal of the American Statistical Association* 91(433): 349-356.

<접수 2020.03.31; 수정 2020.05.26; 게재확정 2020.06.08>

Use of Dual Frame Approach to Sample Supplement and Weighting for Workplace Panel Surveys

Inho Park
Kimin Kim
Jungwoo Kim

In a panel survey, a sequence of interviews are carried out over time to the same sample units that are selected to represent the population of a certain point in time. Collected data from the panel survey enables analyses of the gross changes of characteristics at the unit-level as well as the net change of means at the population level over time. However, it is not possible to avoid the panel attribution, such as non-response, loss of eligibility, and tracking failure, which are caused by continuous follow-up. In particular, panel surveys for establishments are subject to some extreme changes in composition such as closed business and merger, creation of new business sites, changes in the size of businesses such as employees or sales, and changes in business types and regional relocations.

When analyzing the characteristics of the cross-sectional population at the time of tracking using panel data, sample supplement is often considered to reinforce representativeness due to panel attribution. This study examines the dual-frame approach used to supplement samples considered in the 6th wave in 2016 in the Workplace Panel Survey(WPS) conducted by the Korea Labor Institute. The dual frame approach is a sampling technique that utilizes two sampling frames with different coverages in the sample design mainly for cross-sectional population analysis. However, in the case of an establishment panel whose population composition fluctuates over time, the dual-frame approach can be used to supplement a sample that reflects the population change more timely under limited circumstances where there is no business registration. In addition, if the pseudo-maximum likelihood estimation is applied, appropriate weights can be calculated not only for the existing cross-sectional sample but also for the cross-sectional sample expanded by supplementing the sample.

Key words: Workplace Panel Surveys(WPS), dual frame, supplement sample, panel weight, cross-sectional weight