

사업체패널 Small Dataset 소개

김 기 민*

I. 머리말

한국노동연구원에서는 2002년부터 2004년까지 1차 예비조사를 포함하여 3차례에 걸쳐 사업체패널조사를 실시하였다. 사업체패널조사는 국내에서 유일하게 우리나라 사업체의 인적자원관리와 노동수요 및 노사관계 등의 이슈에 대한 대규모 패널데이터를 제공하였다. 이 조사에 대해 설문분량의 부족에 대한 문제 제기와 함께 체계적인 표본설계의 결핍, 데이터 불일치와 로직(logic)상의 오류 등이 일부 지적된 바 있다.

이에 따라 한국노동연구원에서는 보다 양질의 사업체 관련 데이터를 생산하고 제공하기 위해 2002~2004년 사업체패널조사 결과를 분석하여 문제점을 파악함과 동시에 벤치마킹을 위해 외국의 사업체패널조사 사례를 연구하였다. 2005년에는 노동문제 여러 분야의 전문가들로 구성된 포럼을 개최하여 사업체패널조사의 표본과 설문을 재설계하는 등의 준비과정을 거쳤으며, 이러한 노력을 토대로 2006년에는 새롭게 구성된 1,900여 개 표본사업체를 대상으로 「2005 사업체패널조사」를 실시한 바 있다.

「2005 사업체패널조사」가 「2002 사업체패널조사」와 「2003 사업체패널조사」와는 다른 새로운 표본과 설문으로 구성하게 됨에 따라 「2005 사업체패널조사」는 이전 조사인 「2002 사업체패널조사」 및 「2003 사업체패널조사」와는 패널로 연결되지 못하게 되었다. 그러나 많은 연구자들이 사업체를 대상으로 하는 패널데이터의 필요성을 제기함에 따라 패널데이터셋을 구축할 수 있는 방법에 대해 논의하게 되었다. 그 결과 2002년부터 2005년까지의 3개년도 데이터셋의 표본을 구축하여 동일사업체를 추려서 작게나마 패널자료를 구축할 수 있게 되었다.

* 한국노동연구원 책임연구원(kimin1104@kli.re.kr).

이 글에서는 어떠한 방법을 통해 동일사업체를 찾아서 연결했는지, 그리고 그렇게 구축된 데이터의 산업 및 규모가 어떠한지를 살펴보고, 또한 「2005 사업체패널조사」의 설문과 이전 조사의 설문 중 어떤 내용들이 연결되는지를 살펴보도록 하겠다.

II. 사업체패널 Small Dataset

1. 사업체패널 Small Dataset의 결합방법 및 구성 사항

여기서는 사업체패널 Small Dataset을 어떻게 결합하였고, 그 구성사항은 어떠한지 살펴보고자 한다. 2002년부터 2005년까지의 3개 데이터셋에서 같은 사업체가 있는지를 알아보기 위해서는 사업체를 확인할 수 있는 유일한 정보가 필요하다. 우선 각 데이터셋 표본의 사업체명과 주소를 가지고 고용보험 DB를 검색하여 ‘고용보험 사업장관리번호’를 일일이 찾는 작업을 하였다. 이를 토대로 3개의 데이터셋을 결합하였더니 그 중 300여 개의 사업체가 연결되었다. 이 중에 ‘고용보험 사업장관리번호’는 같으나 기본정보가 틀리거나 규모의 변화가 이유 없이 크거나 업종이 상이하게 바뀐 사업체는 제외하였다. 이 때 결합기준은 「2005 사업체패널조사」를 기준으로 연결하였다. 즉, 「2005 사업체패널조사」에 있는 사업체에 「2002 사업체패널조사」 혹은 「2003 사업체패널조사」에 있는 사업체를 결합하는 방식을 취했다.

연결된 사업체수는 <표 1>과 같다. 이러한 작업을 통해 최종적으로 256개 사업체패널 Small Dataset이 구축되었다. 그리고 재무정보에 관한 데이터는 한국신용평가정보에서 제공하는 2001년부터 2005년까지의 기업의 대차대조표 및 손익계산서의 정보로 구축되었다.

사업체패널 Small Dataset은 「2005 사업체패널조사」, 「2003 사업체패널조사」, 「2002 사업체패널조사」와 재무데이터셋 등 총 4개의 데이터셋으로 구성되었다. 이 4개의 데이터셋은 ID라는 변수를 통해 결합시킬 수 있다. 이 때 설문문항이 일치하는 변수는 새로운 변수명을 만들어 연도별로 통일된 변수명을 제시하였다. 예를 들어, <표 2>에서 보듯이, 전체 근로자수는 「2005 사업체패널조사」에서는 EP002, 「2003 사업체패널조사」에

<표 1> 연도별 연결된 사업체수 현황

2005	2003	2002
256개	219개	215개

〈표 2〉 사업체패널 Small Dataset 공통변수 리스트

WPS2002		WPS2003		WPS2005		질문 내용
OLD	NEW	OLD	NEW	OLD	NEW	
WPS0210	EMPNUM02	WPS0310	EMPNUM03	EP002	EMPNUM05	전체 근로자수

서는 WPS0310, 「2002 사업체패널조사」에서는 WPS0210이었으나 여기서는 EMPNUM라는 이름으로 통일하였고, 뒤에 연도를 붙여 구분하였다. 그 외 공통변수에 대한 설명은 사업체패널 Small Dataset의 유저가이드에 상세히 나와 있다.

2. 산업·규모별 현황

앞의 방법을 통해 구축된 총 256개의 사업체에 대한 Small Dataset은 표본설계를 통하여 구축된 것이 아니기 때문에 어떤 특성이 반영된 사업체들이 포함된 것인지 알 수가 없다. 여기서는 사업체조사에서 중요한 기준 변수인 산업과 규모를 살펴 이들 사업체의 특성을 알아보려고 한다. 이 때 조사 당시의 사업체 현황과 비교해 보기 위한 기준으로 사업체기초통계조사를 이용하였다. 산업분류는 2005년의 표본추출 당시의 기준으로 정리하였다.

<표 3>은 「2005 사업체패널조사」의 표본추출 당시의 산업기준으로 이전 3개년 동안의 사업체패널조사 및 사업체기초통계조사¹⁾의 산업별 사업체 분포를 나타낸 것이다.

사업체패널 Small Dataset의 산업 분포는 2002년부터 2005년까지 유사한 분포를 보인다. 2002년의 사업체의 산업 분포는 다른 해에 비해 ‘제조업’의 비중은 높고, ‘사회서비스업’의 비중은 낮게 나타나고 있다. 사업체기초통계조사와 비교하여 보면, 사업체패널 Small Dataset에서는 금융보험업이 10% 이상의 분포를 보여 약 4% 정도인 사업체기초통계조사보다 비중이 높게 나타났다. 반면, 사회서비스업은 사업체패널 Small Dataset에서는 10%의 비중을 보여 사업체기초통계조사의 약 25% 정도의 비중에 비해 크게 낮았다. 이와 같은 원인은 우선 금융보험업의 경우 사업체수가 적은 반면에 조사의 성공률이 높아 결합된 비중이 높게 나타난 것으로 판단되며, 사회서비스업의 경우 사업체패널조사에서는 공공부문을 기관단위로 조사하였으나 사업체기초통계조사는 분사·지사·지점 등의 사업장 단위로 조사를 하였기 때문에 그 비중이 낮게 나온 것으로 판단된다. 그러나 전체적으로 보아 산업별 분포는 사업체기초통계조사와 비교해 큰 차이를 나타내지는 않아 산업에 대해서는 어느 정도 신뢰성을 확보했다고 판단된다.

1) 사업체기초통계조사는 상용근로자 30인 이상이며, 농림어업 및 공공행정, 국방, 사회보장행정산업을 제외하고 분석하였다.

〈표 3〉 사업체패널 Small Dataset의 산업 분포

	2005				2003				2002			
	사업체패널 조사		사업체기초 통계조사		사업체패널 조사		사업체기초 통계조사		사업체패널 조사		사업체기초 통계조사	
	빈도	비율	빈도	비율	빈도	비율	빈도	비율	빈도	비율	빈도	비율
전 체	256	(100.0)	46,780	(100.0)	219	(100.0)	44,211	(100.0)	215	(100.0)	43,339	(100.0)
제조업	111	(43.4)	17,140	(36.6)	90	(41.1)	17,019	(38.5)	111	(51.6)	16,666	(38.5)
전기가스수도업	8	(3.1)	518	(1.1)	5	(2.3)	495	(1.1)	8	(3.7)	485	(1.1)
건설업	5	(2.0)	1,869	(4.0)	5	(2.3)	1,615	(3.7)	4	(1.9)	1,502	(3.5)
개인서비스업	14	(5.5)	3,616	(7.7)	11	(5.0)	3,377	(7.6)	11	(5.1)	3,492	(8.1)
운수업	20	(7.8)	3,242	(6.9)	16	(7.3)	3,139	(7.1)	19	(8.8)	3,167	(7.3)
통신업	4	(1.6)	760	(1.6)	4	(1.8)	751	(1.7)	1	(0.5)	807	(1.9)
금융보험업	29	(11.3)	1,630	(3.5)	25	(11.4)	1,610	(3.6)	23	(10.7)	1,705	(3.9)
사업서비스업 기타	35	(13.7)	5,996	(12.8)	33	(15.1)	5,141	(11.6)	27	(12.6)	4,735	(10.9)
사회서비스업	30	(11.7)	12,009	(25.7)	30	(13.7)	11,064	(25.0)	11	(5.1)	10,780	(24.9)

〈표 4〉는 「2005 사업체패널조사」의 표본추출 당시의 규모기준²⁾으로 이전 3개년의 사업체패널조사 및 사업체기초통계조사의 규모별 사업체 분포를 나타내고 있다.

사업체패널 Small Dataset의 규모 분포를 보면 2005년은 이전의 해에 비해 대규모 사업체의 비중이 줄어들고 소규모 사업체의 비중이 늘어난 것으로 보인다. 그러나 사업체패널 Small Dataset과 사업체기초통계조사를 비교하여 보면, 사업체기초통계조사는 3개년 모두 30~99인의 규모의 사업체 비중이 80% 이상으로 나타난 반면, 사업체패널 Small Dataset은 2002년에는 약 7%, 2003년에는 약 17%이고 2005년에는 약 26%로 낮은 비중을 보이고 있다. 반대로 500인 이상 대규모 사업체의 비중은 사업체기초통계조사에서는 약 1.8% 정도로 나타나지만 사업체패널 Small Dataset에서는 2002년에는 약 41%, 2003년에는 약 34%이고 2005년에는 약 24%로 높은 비중을 보였다. 이는 체계적인 표본추출을 통해 얻어진 데이터가 아니라는 한계를 드러내는 지점이다. 그러나 2002년부터 2005년까지 대규모 사업체의 변화에 대해 연구는 충분히 수행할 수 있을 것으로 기대한다. 더불어 대기업뿐 아니라 2002년부터 2005년까지의 사업체 전반의 변화를 분석하기 위해서는 사후가중치를 통한 규모에 대한 보정이 반드시 필요하다고 판단된다.

2) 규모는 30~99인, 100~299인, 300~499인, 500인 이상 4가지 기준으로 분류하였다.

〈표 4〉 사업체패널 Small Dataset의 규모 분포

	2005				2003				2002			
	사업체패널 조사		사업체기초 통계조사		사업체패널 조사		사업체기초 통계조사		사업체패널 조사		사업체기초 통계조사	
	빈도	비율	빈도	비율	빈도	비율	빈도	비율	빈도	비율	빈도	비율
전 체	215	(100.0)	43,339	(100.0)	219	(100.0)	44,211	(100.0)	256	(100.0)	46,780	(100.0)
30~99인	55	(25.6)	34,841	(80.4)	38	(17.4)	35,580	(80.5)	19	(7.4)	37,958	(81.1)
100~299인	59	(27.4)	6,793	(15.7)	51	(23.3)	6,937	(15.7)	64	(25.0)	7,226	(15.4)
300~499인	49	(22.8)	909	(2.1)	55	(25.1)	914	(2.1)	69	(27.0)	807	(1.7)
500인 이상	52	(24.2)	796	(1.8)	75	(34.2)	780	(1.8)	104	(40.6)	789	(1.7)

3. 설문 내용

「2005 사업체패널조사」의 설문은 「2002 사업체패널조사」나 「2003 사업체패널조사」의 설문과 달리 정성적인 문항을 많이 배제하고 기초적인 사실 조사에 중점을 두어 설계하였다. 이로 인해 그 이전 조사와의 공통되는 설문 문항에는 한계가 있을 수밖에 없다. 여기서는 「2005 사업체패널조사」의 설문을 「2002 사업체패널조사」 및 「2003 사업체패널조사」의 설문과 비교하여 공통적으로 조사된 설문 문항에 대해 살펴보고자 한다.

근로자 현황 설문을 살펴보면, 전체 근로자수, 장애인 근로자수와 일부 비정규직 근로자(단시간근로자, 파견근로자, 용역근로자, 일시대체근로자)수에 대해 공통적으로 묻고 있다. 이 때 「2005 사업체패널조사」에서는 파견/용역/일시대체근로자수를 모두 합해 묻고 있고, 「2002 사업체패널조사」 및 「2003 사업체패널조사」에서는 이들 각각의 근로자수를 조사하였다.

다음으로 인사담당자 설문 중 각 부분별로 공통적으로 조사된 문항에 대해 살펴보겠다.

사업장 특성 부분에서는 기본사항인 분사 여부와 사업장수³⁾, 기업에 관련된 정보인 경영체제, 외국인 지분율, 지분을 가진 외국인의 역할, 주력사업에서의 해외진출에 관한 전략 유무, 사업장에 관련된 정보인 사업장의 설립연도, 주된 제품/상품이나 서비스의 가격·품질·전략에 대한 경쟁업체와의 수준 비교, 하도급 거래형태에 대한 설문 등이 공통적으로 나타났다.

고용현황 및 관리 부분에서는 채용수단과 관련항목에 대해서는 모두 조사하였다. 이 때 「2005 사업체패널조사」는 우선순위를 두지 않고 복수응답 형식으로 질문하였고, 과

3) 「2005 사업체패널조사」에서는 국내·국외로 나눠서 조사하였다.

거 설문은 가장 많이 사용하는 1, 2순위를 묻는 방식으로 조사하였으며, 발탁승진제도에 대한 설문은 「2005 사업체패널조사」에서는 제도도입 유무만을 조사하였고, 과거 설문에서는 발탁승진자의 비율에 대해서도 조사하였다. 또한 정규직 근로자에 대해 경영상의 이유로 인한 해고를 하지 않는다는 경영정책이나 합의사항 여부에 대한 공통설문이 있었고, 비정규근로자와 관련해서는 고용추세와 사회보험 및 부가급부에 대한 적용률을 공통적으로 조사하였다.

보상 및 평가 부분에서는 인사사고 및 근무평정과 관련하여 다면평가, 목표설정(MBO)방식 활용과 인사사고 결과 공개 여부, 임금체계와 관련하여 연봉제의 실시 여부 및 적용대상자에 대해 공통적으로 조사되었다. 또한 성과배분과 관련해서 성과배부 실시 여부, 우위사주제의 도입 여부 및 지분율, 스톡옵션제 도입 여부 및 적용대상에 대해, 임금수준과 관련해서는 군필 대졸 남성 정규직 신입사원 평균 연봉과 국내 동종업계와 비교시 임금수준에 대해 공통적으로 조사되었다.

인적자원관리 및 작업조직 부분에서는 인사관련 전반적인 상황과 관련하여 인사·노무 담당자의 구성현황, 핵심인력 확보 및 육성, 유지프로그램, 인사관리의 특성에 대한 주관적인 인식 척도(인사관리의 일차적 목표, 인력 유지 및 운영방안, 인사관리 기준 등)가 5점 척도로 공통적으로 조사되었다. 또한 인사관련 이슈와 인사책임자의 역할 등에 대해 공통적으로 조사되었으며, 작업조직과 관련해서 식스 시그마 도입 여부, 소집단 활동에 관한 사항, 업무 로테이션 관련 사항, 다기능 교육훈련에 관한 사항, 작업단위의 자율성(업무수행 방법, 업무속도, 신규팀원 채용, 팀원 훈련) 등이 공통적으로 4점 척도 방식으로 조사되었다. 근로시간과 관련해서는 주 5일제, 교대근무제 실시 여부 및 교대제 형태와 선택적·탄력적 근무시간제 운영 여부에 대해 공통적으로 조사되었으며, 공정한 처우 및 모성보호 관련해서는 육아휴직, 배우자 출산휴가(남성), 직장보육시설, 생리휴가, 수유시간 보장, 임신한 여성의 야간근무·휴일근무·초과근무 제한, 임신한 여성의 유해업종 종사금지 여부, 임신한 여성에게 작업전환 시행 여부, 출산 후 1년이 되지 않은 여성에게 유해업무 종사금지 여부, 태아검진휴가와 유·사산 휴가 여부에 대한 항목들이 공통적으로 조사되었다. 특히 인적자원관리 및 작업조직 분야에서 공통적으로 묻고 있는 항목이 많은 것으로 나타났다. 따라서 인사관리 현황, 작업조직 현황이나 모성보호관련 제도에 대해서는 좀더 심도 깊은 연구를 할 수 있을 것이라 판단된다.

인적자원개발 부분에서는 교육훈련을 실시 혹은 지원 여부와 교육훈련 전담부서와 전담자에 대한 사항만이 공통적으로 조사되었다.

기업복지 및 산업재해 부분에서는 퇴직금제도의 형식, 퇴직금제도의 퇴직연금제도로 전환할 가능성과 전환시의 형태와 선택적 복지제도 운영 여부에 대한 사항이 공통적으로 조사되었다.

노사관계 설문을 보면, 노동조합 유무, 노사관계의 전반적인 상태, 노동조합의 현황, 기업단위 임금교섭의 구조, 임금인상에 대한 사항, 파업에 관련된 사항과 노사협의회 유무에 대한 정보들이 공통적으로 들어 있다.

마지막으로 재무정보는 한국신용평가정보에서 제공하는 2001년부터 2005년까지의 기업의 대차대조표 및 손익계산서 자료, 재무비율 정보를 제공하고 있다. 대차대조표 정보에는 자산, 부채와 자본에 대한 내용이 들어 있고, 손익계산서 정보에는 매출액, 매출원가, 판매비와 관리비(인건비, 복리후생비, 교육훈련비, 광고선전비, 연개개발비 등), 영업이익, 영업외수익, 당기순이익 등에 대한 내용이 들어 있다.

III. 맺음말

보다 체계적인 조사를 위하여 새로운 표본과 설문을 설계함에 따라 「2002 사업체패널조사」와 「2003 사업체패널조사」의 패널데이터를 살리지 못한 채 「2005 사업체패널조사」를 조사하였다. 이에 따라 2002년부터 2005년까지 세 차례 사업체조사를 수행했음에도 불구하고, 그동안 본격적인 패널분석 방법론을 사용해 보지 못하고, 다만 횡단면 분석에만 의존할 수밖에 없었다. 많은 연구자들이 사업체를 대상으로하는 패널자료에 대한 필요성을 제기함에 따라 어느 정도의 패널분석을 할 수 있도록 사업체패널 Small Dataset을 구축하였다. 또한 한국신용평가정보의 2000년부터 2005년까지의 대차대조표 및 손익계산서의 내용을 바탕으로 재무정보를 구축하였다.

2002년부터 2005년까지 3개의 사업체조사 데이터셋과 1개의 재무 데이터셋으로 총 4개의 데이터셋을 구성하였고, 이 데이터를 ID변수를 통해 결합하였으며, 공통된 설문 문항에 대해서는 통일된 변수명을 제공하였다.

최종 결합된 총 256개 사업체의 산업 및 규모의 분포를 살펴본 결과, 산업의 분포는 사업체기초통계조사와 비교하였을 때 어느 정도 신뢰성을 확보하는 것으로 보이나, 규모의 분포는 사업체패널 Small Dataset에서 대규모 사업체의 비중이 매우 높게 나타나 그 당시의 정확한 전체 사업체 분포를 반영하지는 못하는 것으로 보인다. 그러나 추후 사후가중치 부과 등의 방법을 통해 이러한 결점을 다소나마 보정할 수 있을 것으로 기대한다.

설문내용상의 연속성 측면을 보면, 공통된 설문을 일정수준 이상 확보하고 있다. 특히 인적자원관리 및 작업조직의 설문은 다른 설문에 비해 많은 설문이 일치하고 있어 이 부분에 대해서는 보다 심도 깊은 연구를 할 수 있을 것이라 판단된다.

한국노동연구원에서는 사업체패널 **Small Dataset**을 가지고 분석한 논문으로 2008년 10월에 워크숍을 개최할 예정이다. 이 워크숍에서 의미있는 여러 편의 논문들을 통해 좀 더 나은 패널데이터 구축의 계기가 될 수 있기를 기대한다. **KLI**