

2012년도 국가통계 품질개선 컨설팅 연구용역

『환경보호지출계정』
품질개선 컨설팅 최종결과보고서

2012. 8.

제 출 문

제 출 문

통계청장 귀하

본 보고서를 「환경보호지출계정」 품질개선 컨설팅의 최종
보고서로 제출합니다.

2012년 8월

한국조사연구학회 ㉠

연구진

책임연구원	이 계 오, 한남대학교 비즈니스통계학과 교수
연구원	이 기 재, 방송통신대학교 교수
연구보조원	홍 영 택, 한국갤럽조사연구소 연구원
통 계 청	김 경 용, 통계청 품질관리과 사무관 윤 은 경, 통계청 품질관리과 주무관

최종결과보고서 요약문

연구과제명	「환경보호지출계정」 품질개선 컨설팅
주 제 어	환경보호지출계정, 표본추출틀, 절사층화추출법
연 구 기 간	2012년 6월1일 ~ 2012년 8월24일
연 구 기 관	한국조사연구학회
연구진구성	책임연구원 : 이계오(한남대학교 교수) 공동연구원 : 이기재(방송통신대학교 교수), 김경용(통계청 품질관리과 사무관) 윤은경(통계청 품질관리과 주무관) 연구보조원 : 홍영택(한국객립조사연구소 연구원)
<p>「환경보호지출계정」의 민간기업부문조사와 관련하여 2010년 정기품질진단에서 제기된 문제점인 제조업과 비제조업간의 조사단위 불일치문제와 표본의 최신화에 대한 개선방안 연구와 신규표본설계의 표본추출 가이드라인을 제시하는 것이 주요연구내용이며 이에 대한 연구결과를 아래와 같이 요약하였다.</p> <p>먼저 제조업과 비제조업의 조사단위 불일치는 2010년 경제총조사의 데이터를 조사모집단으로 사용할 것으로 제안함으로써 해결할 수 있다.</p> <p>두 번째는 모집단 층화에서 업종구분을 제9차 표준산업분류체계(KSIC 9)를 사용함으로써 기존 통계와 시계열 유지를 위한 제8차 표준산업분류체계(KSIC 8)와 KSIC 9와 연계하는 방안을 제시하였다.</p> <p>마지막으로는 신규표본설계의 표본추출가이드라인으로 ① 절사층화추출법으로 Lavallée와 Hidiroglou 방법의 적용절차의 설명, ② 층화변수의 선정, ③업종별 적정한 표본크기 계산(목표허용오차를 이용한 산식), ④ 층별 표본배분과 표본추출방안, ⑤ 표본가중치 산출과 모수 추정법, ⑥ 특이치 검출방법 및 ⑦ 모집단 변동과 표본사업체 관리 등에 대한 절차를 제시하였다.</p>	

차 례

제 1 장 연구배경	1
제 2 장 통계작성 실태분석	2
1. 통계 작성 개요	2
2. 표본추출방법	2
3. 가중치 및 모수추정	3
제 3 장 통계 품질 개선 내용	6
1. 표본추출틀 변경 방안	6
2. 표본설계 개선방안	6
3. 기존 환경보호지출계정과 개선안간의 시계열유지 방안	8
제 4 장 업종분류체계 변경	10
1. 변경내용	10
2. 일관성 확보 방안	11
제 5 장 표본추출틀 변경	13
1. 「전국사업체조사」 개요	14
2. 「전국사업체조사」 결과 분석	16
3. 일관성 확보 방안	23
제 6 장 최적표본설계방안 : 신규표본설계의 표본추출 가이드라인	24
1. 표본설계의 기본원칙	24
2. 변형절사추출법의 적용 방안	25
3. 층화변수에 대한 검토	26

4. 표본크기 결정의 원칙	26
5. 특이치 검출 방안	28
6. 가중치 작성 및 추정방법	31
7. 2013년 조사를 위한 표본설계 연구절차	32
제 7 장 결언	42
[부록 1] 다양한 변동계수 크기에 따른 표본크기 산출결과	44
<참고문헌>	45

표 차 례

<표 1> 제조업의 일부업종의 모집단규모와 표본규모 및 가중치 산출결과	5
<표 2> 비제조업 일부 업종의 모집단규모와 표본규모 및 가중치 산출결과	5
<표 3> 2009년 조사 데이터의 인건비와 환경보호경상지출의 분석결과	7
<표 4> KSIC 8과 KSIC 9의 산업대분류 연계표	9
<표 5> 표본사업체의 관리번호	10
<표 6> 제8차 산업분류와 제9차 산업분류의 연계표	11
<표 7> 산업대분류별 사업체 규모별 모집단 사업체 수 현황	19
<표 8> 산업대분류별 사업체 규모별 모집단 근로자 수 현황	20
<표 9> 업종 구분별 사업체 규모별 모집단 사업체 수 현황	21
<표 10> 업종 구분별 사업체 규모별 모집단 근로자 수 현황	22
<표 11> 표본개선안의 조사모집단과 필요 표본크기와 기대변동계수	28
<표 12> 음식료 제조업(담배제조업 포함)에 대한 층 경계점 결정	38
<표 13> 섬유, 의복, 모피 제조업에 대한 층 경계점 결정	39

제 1 장 연구배경

- 환경보호지출계정은 환경정책평가연구원에서 경제주체들이 환경보호를 위해 실제로 지출하거나 부담하는 비용을 환경영역별로 체계적인 정리를 통해서 생산하는 통계이다.
- 본 통계는 환경정책의 유효성 및 효율성 분석, 환경시장 규모 파악 등에 필요한 통계로, 많은 선진국들이 작성하고 있을 뿐아니라 OECD에도 제출해야하는 환경분야의 대표적인 통계 중 하나이며 국가승인통계로 작성하고 있다.
- 국가승인통계의 통계품질의 실태파악과 품질제고 방안을 마련하기 위해서 품질진단을 실시하고 있다. 2010년도 ‘환경보호지출계정’의 통계품질진단결과에서 도출된 문제점을 보완하기 위한 개선방안을 연구하는데 목적이 있다.
- 2010년도 정기통계품질진단에서 문제점으로 지적된 중요한 사항은 아래 두 가지이다.
 - 제조업 및 비제조업간의 조사단위 불일치 문제 해소 필요
그동안 제조업은 사업체 단위로, 비제조업은 기업체 단위로 조사함으로써 조사단위가 불일치하여 다양한 업종의 사업체로 구성된 기업체에서 환경보호지출계정을 정확하게 파악하기가 곤란하였으므로 이의 개선이 필요함
 - 환경보호지출계정 표본의 최신화 필요
표본 추출틀로 사용한 제조업은 ‘03년 기준 광업·제조업조사, 비제조업은 ‘03기준 국세청 법인세 신고업체들이므로 신규 기업의 진입과 기존기업의 퇴출을 반영하고 ‘09년에 개정된 산업분류체계의 적용과 이에 따른 시계열 유지에 대한 연구가 필요함
- 2010년 품질진단의 결과보고서를 검토하여 문제점으로 지적되지는 않았지만 환경보호지출계정의 표본조사에서 사용되는 가중치 산출과 모수추정방법의 개선방안 및 비표본오차의 발생요인들에 대한 분석과 개선안의 연구가 필요함

제 2 장 통계작성 실태분석

2.1 통계 작성 개요

- 「환경보호지출계정」은 통계청 승인 가공통계로서 환경부에서 한국환경정책·평가연구원에 위탁하여 생산하는 통계이다. 작성연혁은 1994년에 환경보호지출계정을 편제하기 위한 준비작업 일환으로 연구를 시작하였으며 1997년 통계청의 승인을 득하여 한국은행에서 통계를 생산하였고, 2003년에는 OECD의 환경관련 요구 통계의 변경으로 환경보호지출과 수입통계를 통합하여 작성한 후에 발표하였다. 2009년에는 통계작성기관이 한국은행에서 환경부로 변경되었으며 경제와 환경을 통합하는 정책수단의 도입 및 환경정책의 체계적이고 효과적인 평가를 위해서 환경경제통합계정의 하위계정인 환경보호지출계정을 개발하여 공표하고 있다.
- 조사목적은 환경보호지출계정은 경제활동과정에서 발생하는 환경오염을 예방하고 오염된 환경을 복구하는데 사용된 경제자원을 평가하기 위해 환경보호서비스 생산 및 환경보호지출을 경제주체별, 환경영역별로 체계적으로 분류한 통계이며, 국가 전체의 환경보호지출 규모 및 자금부담 내역, 환경보호서비스의 생산과 사용 등을 파악할 수 있으며, 환경요인조정 국민순생산(EDP : Environmentally-adjusted net Gross Domestic Product)의 추계에 필요한 환경계정편제의 기초통계로 활용하고자 한다.
- 조사대상은 4개 부문으로 정부부문, 기업부문, 가계부문과 환경전문업 등이다. 환경보호지출계정에서 핵심적으로 다룰 부문은 기업부문으로 우리나라 기업 중 제조업(총생산액) 및 비제조업(총매출액)의 90% 사업장(업체) 등이다. 앞으로는 특별한 언급이 없는 한 기업부문 관한 조사통계부문을 지칭한다.

2.2 표본추출방법

- 조사범위는 우리나라 전 산업을 조사하는 것을 원칙으로 하나 환경오염방지지출 규모가 미미할 것으로 보이는 재배업, 수렵업, 임업, 도소매업 중 자동차판매 및 수선업, 금융보험부동산 및 사업서비스업, 사회 및 개인서비스업중 보건업 등은 제외하였다.
- 산업분류는 한국표준산업분류(KSIC)에 따르며 대분류단위는 OECD의 질문지(Questionnaire)를 참고로 한다.

- 모집단 및 표본프레임에서 모집단 구성은 조사대상업종의 모든 기업체로 구성하는 것이 바람직하지만 자료이용의 제약으로 국세청 법인세 신고대상업체 중에서 법인기업체를 표본프레임으로 정하였으며 1992년도 법인세 신고기업 94,883개 중에서 비조사대상 업종의 기업과 매출액이 없는 기업 등을 제외한 49,554개 기업이 표본프레임이다.
「광공업통계조사」 보고서에 따르면 전체 기업체에 대한 법인기업체의 비중은 매출액 기준 89.6%, 유형고정자산 기준 87.4%이므로 모집단과 표본프레임의 차이에 따른 오차는 법인기업 가중치를 적용하여 추계함으로써 전체 모집단에 관한 환경지출을 추계할 수 있을 것이다.
- 조사대상업체 선정은 일종의 질사층화추출법을 적용하여 전수조사업체와 표본조사업체 구분하여 선정하였는데 전수조사업체는 환경오염산업 중 매출액 500억 원 이상, 여타 산업 중 매출액 1000억 원 이상 업체와 업종별 매출액 상위 5대 업체로 하였으며 환경오염산업은 산업은행의 공해방지투자규모 및 환경부의 「제조업체 환경오염방지활동 및 투자현황」을 근거로 8개 업종(종이제조업, 석유정제품, 화합물·화학제품, 고무·플라스틱, 비금속광물, 제1차 금속, 조립금속 및 기계장비제조업)을 선정하였다.
표본조사업체의 선정은 전수조사업체를 제외한 기업체들을 환경투자 및 비용과 상관관계가 높은 것으로 조사된 매출액을 대표변량으로 하여 업종별로 1차 층화하고 매출액 규모로 2차 층화한 후에 네이만배분법을 적용하여 층별로 표본크기를 산출하였으며 각 층에서 랜덤하게 표본업체를 선정한 일종의 층화확률추출법을 적용하여 표본사업체를 선정하였다.
- 전체 표본크기는 제조업 1,746개 업체이고 비제조업 1,224개 업체로 2,970개 업체이다.

2.3 가중치 및 모수추정

- 가중치 산출은 전수조사층과 표본조사층을 구분하여 산출하였는데 전수조사층에서는 생산액이나 매출액을 기준으로 2003년을 기준연도로 정하고 조사해당년도의 매출액의 비를 가중치로 계산하였다. 아래 표로 주어진 제조업과 비제조업에서 전수조사층에는 각 층별로 2003년 기준 생산액과 조사해당년도의 생산액(매출액)에 주어졌으며 조사해당년도의 생산액을 2003년 생산액으로 나누는 값을 해당 층의 가중치로 주어졌다. 예를 들어 음식료업종의 전수조사층은 68개 사업체가 조사대상 이었고 조사 회수된 사업체는 56개 사업체였으며 이들 생산액의 2003년 기준액과 2009년 기준액은 각각 14,046,593만원과 16,678,963만원이었으므로 이들의 비인 $1.1874(=16,678,962/14,046,593)$ 를

가중치로 사용하였다.

- 표본조사층에서 가중치는 모집단의 사업체 수를 조사 완료된 사업체 수로 나누어서 계산하였다. 예를 들어 음식료업의 1층에서 모집단 사업체 수는 141개이고 조사 완료된 사업체 수는 53개이므로 산출한 가중치는 2.6604(=141/53)이다.
- 표본조사층과 전수조사층이 다른 기준으로 가중치를 산출하였으며 가중치 적용의 본래 취지와 상이한 가중치 산출과정을 적용하였으므로 모수의 참값을 추정할 수 있는지에 대한 정확한 근거를 찾을 수 없으므로 가중치는 표본추출률의 역수와 응답률의 역수를 곱해서 설계가중치를 산출한 다음에 사후층화보정가중치를 산출하여 설계가중치에 곱해서 최종가중치를 산출하는 방법으로 개선해야할 것이다.
- 모평균에 대한 모수추정량은 표본가중평균으로 아래 식으로 계산할 수 있다.

$$\hat{\mu}_h = \frac{\sum_{i=1}^{n_{hi}} \sum_{j=1}^{n_{hi}} W_{hij} y_{hij}}{\sum_{i=1}^{n_{hi}} \sum_{j=1}^{n_{hi}} W_{hij}}$$

여기서 h 는 업종을 나타내고 i 는 규모층을 나타내며 j 는 표본사업체를 의미하고 W_{hij} 와 y_{hij} 는 각각 h 업종 i 규모층 내의 j 사업체의 가중치와 관찰 값을 나타낸다.

<표 1 > 제조업의 일부업종의 모집단규모와 표본규모 및 가중치 산출결과

업종층	조사 모집단수	생산액 (매출액)	표본수	회수	생산액 (2003년)	가중치	업종명
1	68	16,678,962	68	56	14,046,593	1.1874	음식료
2	141		24	53		2.6604	음식료
3	380		24	32		11.8750	음식료
5	47	6,539,724	47	29	4,187,671	1.5617	섬유의복모피
6	209		32	39		5.3590	섬유의복모피
7	784		30	36		21.7778	섬유의복모피
9	39	1,385,888	39	28	1,039,796	1.3328	목재나무
10	216		25	15		14.4000	목재나무
11	44	6,625,206	44	34	5,210,096	1.2716	펄프종이
12	100		24	36		2.7778	펄프종이
14	62	5,060,727	62	41	3,273,741	1.5459	출판인쇄
15	157		26	32		4.9063	출판인쇄
16	364		20	24		15.1667	출판인쇄

<표 2> 비제조업 일부 업종의 모집단규모와 표본규모 및 가중치 산출결과

업종층	조사 모집단수	생산액 (매출액)	표본수	회수	생산액 (2003년)	가중치	업종명
71	74	6,044,660	74	54	2,998,383	2.0160	음식숙박
72	267		22	24		11.1250	음식숙박
73	41	37,159,435	41	25	24,093,632	1.5423	운수창고
74	121		19	19		6.3684	운수창고
75	502		20	20		25.1000	운수창고
77	14	32,021,653	14	11	28,246,098	1.1337	통신
79	98	257,521,260	98	75	210,356,075	1.2242	금융부동산임대
80	455		22	23		19.7826	금융부동산임대
82	77	8,939,325	77	59	6,848,843	1.3052	보건
83	299		29	43		6.9535	보건
84	58	31,413,584	58	44	21,396,328	1.4682	기타서비스업
85	187		30	31		6.0323	기타서비스업
86	442		20	20		22.1000	기타서비스업

제 3 장 통계 품질 개선 내용

3.1 표본추출틀 변경 방안

- 기존 민간기업 부문 조사의 표본설계에서는 제조업의 추출틀은 통계청의 광업·제조업조사의 조사단위인 사업장을 조사단위로 사용하였고 비제조업의 추출틀은 국세청의 법인세 신고 기업 중에서 매출액 정보가 있는 기업을 사용하였다. 이와 같은 방식은 사업장과 기업체의 추출단위의 포괄범위가 상이함으로 인한 포괄범위의 비표본오차가 발생할 수 있으므로 동일한 수준의 추출단위를 추출틀로 사용하는 방안이 연구되어야 할 것이다. 또한 광업·제조업조사는 종사자수 10인 이상인 사업체만을 조사대상으로 하고 있으므로 전체 사업체를 포함하고 있지 않기 때문에 전체 총계에 대한 모수추정에서는 축소편향이 발생할 가능성이 있다.

최종적으로 생산액이나 매출액의 정보와 종사자 수의 정보를 포함한 경제총조사의 데이터를 표본추출틀로 사용할 것을 제안하지만 자료분석의 제약으로 본 연구에서는 현재 시점에서 활용할 수 있는 최신의 표본추출틀 정보인 2010년 12월 말 기준의 「전국사업체조사」의 데이터를 이용할 수밖에 없으나 기본적인 표본추출방법에 대한 연구 내용은 동일한 것이므로 문제가 되지 않을 것이다.

- 절사층화추출방법을 적용하기 위해서 분석할 변수는 생산액(매출액)을 사용하였으나 2010년 「전국사업체조사」에서는 생산액(매출액)의 정보를 이용할 수 없으므로 상관성이 깊은 종사자 수를 분석변수로 사용하여 전수조사대상 사업체를 결정하는 방법을 연구할 것이다.

3.2 표본설계 개선방안

- 사용할 표본추출틀을 확정된 후에 표본설계 연구에서 중요한 요소는 표본크기의 계산인데 표본크기는 일반적으로 가용한 조사비용의 금액 또는 추정치의 목표표본오차의 크기에 의해서 결정하지만 여기서는 2009년의 조사된 데이터 중에서 관심조사항목인 개인당 인건비와 환경보호경상지출금액의 변동계수를 산출하여 목표변동계수를 선정하는 방법으로 표본크기를 검토하겠다.

업종별 규모 층별로 인건비와 환경보호경상지출금액에 대한 추정치와 표준오차와 변동계수의 산출한 결과를 아래 <표 3>에 정리하였다.

- <표 3>에 주어진 정보와 2010년 「전국사업체조사」 자료를 분석하여 표본크기

의 적절성을 검토하기 위해서 업종별로 적절한 기대변동계수의 크기를 선정하는 분석을 할 것이다.

<표 3> 2009년 조사 데이터의 인건비와 환경보호경상지출의 분석결과

층	중분류	소분류	표본 크기	인건비			환경보호경상지출		
				평균	se	CV	평균	se	CV
1-4	음식류		141	115.77	9.66	8.35	402.86	43.87	10.89
5-8	섬유의복모피		104	73.08	10.60	14.51	152.92	32.03	20.95
9-10	목재나무		43	98.68	18.80	19.06	145.73	51.05	35.03
11-13	종이인쇄출판	필프종이	70	462.10	55.25	11.96	2059.57	299.81	14.56
14-16		출판인쇄	97	23.15	6.21	26.81	25.42	4.94	19.44
17-18	석유정제 화학고무	코크스석유정제	26	896.19	332.95	37.15	8117.88	3797.05	46.77
19-21		화합물화학	177	140.61	21.57	15.34	601.39	95.49	15.88
22-25		고무플라스틱	151	60.40	17.83	29.51	111.17	27.86	25.06
26-29	비금속광물		148	58.42	7.52	12.87	177.90	39.41	22.15
30-32	제1차금속		130	97.49	21.48	22.03	1200.75	692.72	57.69
33-37	조립금속	조립금속	139	36.41	4.63	12.71	73.78	11.09	15.03
38-42	기계장비	기계장비	211	78.68	13.66	17.36	219.54	74.31	33.85
43-45	운수장비		99	250.23	49.80	19.90	784.06	172.76	22.03
46-49	가구기타		122	32.24	5.50	17.06	70.35	14.54	20.67
51-52	축산재배임업		61	62.43	6.76	10.82	114.64	28.15	24.55
53-54	어업		16	143.23	18.01	12.58	657.28	141.67	21.55
55-56	광업		32	114.55	76.29	66.60	173.71	88.30	50.83
57-58	전기가스수도		40	2058.11	605.30	29.41	16699.00	4958.36	29.69
59-64	건설업		103	596.79	190.22	31.87	743.66	327.60	44.05
65-70	도소매업		161	61.83	26.53	42.90	75.40	31.41	41.66
71-72	음식숙박		78	244.87	90.67	37.03	277.15	121.82	43.95
73-76	운수창고통신	운수창고	64	99.93	38.41	38.44	423.46	202.11	47.73
77-78		통신	11	.	.	.	16.35	5.95	36.38
79-81	금융부동산 임대		98	144.99	79.07	54.53	986.62	418.62	42.43
82-83	보건		102	67.72	8.52	12.58	146.88	28.96	19.71
84-88	기타서비스업		95	126.81	28.48	22.46	190.01	85.53	45.01

- 표본추출방법의 가이드라인은 모집단 층화방법에서 층의 경계점 계산방법, 전수조사층의 경계점 결정방법, 층별 표본배분법 선택, 표본추출방법 및 선정된 표본사업체의 관리방법과 조사된 데이터의 에디팅절차, 가중치 산출과 모수추정방법 등을 세부적으로 기술하는 내용이 될 것이다.

3.3 기존 환경보호지출계정과 개선안간의 시계열유지 방안

- 기존 환경보호지출계정에서 사용한 업종별 분류는 제8차 산업분류(KSIC 8)이지만 2010년 「전국사업체조사」에 적용된 산업분류는 제9차 산업분류(KSIC 9)이므로 KSIC 8과 KSIC 9를 연계하는 연구가 우선되어야 한다. 우선 산업대분류수준에서 KSIC 8과 KSIC 9를 연계하는 내용을 축약 정리하면 아래 <표 4>와 같다.
- 2009년 환경보호지출계정 조사에서 응답한 사업체들의 산업분류코드에 대한 정보는 중분류 수준까지 있으므로 중분류이상 수준에서는 제8차 산업분류(KSIC 8)에서 제9차 산업분류(KSIC 9)으로 변경사항을 이용하여 시계열유지 방안을 분석할 수 있으나, 중분류이하 차원에서 제8차 산업분류(KSIC 8)에서 제9차 산업분류(KSIC 9)로 변경된 경우에는 연계할 수 있는 방법이 없으므로 시계열유지에서 제한사항이 될 수 있다. 예를 들면 제조업에서 일부는 하수·폐기물 처리업으로, 또 다른 일부는 출판·영상·방송통신 및 정보서비스업으로 변경되었으므로 이에 대한 추적분석은 심층적인 연구가 필요하다.
- 2009년 환경보호지출계정의 자료를 이용하여 KSIC 8과 KSIC 9의 연계표에 의한 모집단 변동에 따른 차이를 분석하여 신규 표본설계에 의한 모수추정 결과의 변동을 보정할 수 있는 보조정보로 사용가능성을 검토한다.
- 기존 환경보호지출계정에서 사용한 가중치 산출방법과 신규 표본설계에 적용한 가중치 산출방법의 차이점을 분석하여 모수추정의 차이에 대한 요인을 파악하고 기존 모수 추정치와 신규 표본설계에 의한 모수 추정치의 차이를 보정하는 보조정보로 활용 가능성을 분석한다.
- 기존 환경보호지출계정의 모수 추정치와 신규 표본설계의 모수 추정치 간의 차이에 대한 정확한 분석을 위해서 신규표본설계에 의해 수집된 조사 자료에 기존 가중치·모수 추정법과 신규 가중치·모수 추정법을 병행 적용하여 차이의 크기를 분석하여 시계열 유지를 위한 보정계수의 적용을 검토한다.

<표 4> KSIC 8과 KSIC 9의 산업대분류 연계표

KSIC 8	KSIC 9
A. 농업 및 임업	A. 농업, 임업 및 어업
B. 어업	
C. 광업	B. 광업
D. 제조업	C. 제조업
E. 전기, 가스 및 수도사업	D. 전기, 가스, 증기 및 수도사업
F. 건설업	E. 하수·폐기물처리, 원료 재생 및 환경복원업
G. 도매 및 소매업	F. 건설업
H. 숙박 및 음식점업	G. 도매 및 소매업
I. 운수업	H. 운수업
J. 통신업	I. 숙박 및 음식점업
K. 금융보험업	J. 출판, 영상, 방송통신 및 정보서비스업
L. 부동산 및 임대업	K. 금융 및 보험업
M. 사업서비스업	L. 부동산업 및 임대업
N. 공공행정, 국방 및 사회보장행정	M. 전문, 과학 및 기술서비스업
O. 교육 서비스업	N. 사업시설관리 및 사업지원서비스업
P. 보건 및 사회복지사업	O. 공공행정, 국방 및 사회보장행정
Q. 오락, 문화 및 운동관련산업	P. 교육서비스업
R. 기타 공공, 수리 및 개인서비스업	Q. 보건업 및 사회복지서비스업
S. 가사서비스업	R. 예술, 스포츠 및 여가관련 서비스업
T. 국제 및 외국기관	S. 협회 및 단체, 수리 및 기타 개인서비스업
	T. 가구내 고용활동 및 달리 분류되지 않은 자가생산활동
	U. 국제 및 외국기관

제 4 장 업종분류체계 변경

- 2008년 통계청에서 제9차 한국표준산업분류를 개편하였으나 2009년 환경보호지출계정에서 적용한 산업분류는 제8차 산업분류이다. 앞으로 표본추출틀로 사용한 데이터는 모두 산업분류가 제9차 분류체계를 적용하였으므로 8차 기준으로 분류된 산업분류를 9차 분류기준으로 전환해야 한다. 그러나 기존의 환경보호지출계정에서 조사된 사업체의 분류는 2자리 코드(중분류)까지만 주어졌으므로 정확하게 산업분류를 전환할 수는 없으나 최대한 주어진 정보를 이용하여 산업분류 체계변경을 설명하겠다.

4.1 변경내용

- 2009년 환경보호지출계정의 데이터에 부여된 관리번호는 원표본과 예비표본을 구분하고 다음에는 2자리 코드로 업종을 구분하였으며 다음 3자리 코드는 표본사업체의 일련번호를 나타내고 있다.
⇒ 모든 표본사업체에 대하여 관리번호를 부여

<표 5> 표본사업체 관리번호

○	△ △	◇ ◇ ◇	① : 원표본 / 예비표본 구분
①	②	③	② : 업종구분(1~49 제조업, 51~88 비제조업) * 전수층 : 1, 5, 9, 11, 14, 17, 19, 22, 26, 30, 33, 38, 43, 46, 51, 53, 55, 57, 59, 65, 71, 73, 77, 79, 82, 84
			③ : 일련번호

“예” 22004 - 원표본(전수층), ‘22층’ 업종의 4번째 기업

725004 - 제조업, 예비표본, ‘25층’ 업종의 4번째 기업

856004 - 비제조업, 예비표본, ‘56층’ 업종의 4번째 기업

- 2009년 환경보호지출계정의 조사에서 업종별로 전수조사층과 표본조사층으로 구분하고 표본조사층은 규모에 따라서 3개 층으로 구분하였는데 구분에 사용한 정확한 기준이 없으므로 새로운 표본추출틀을 적용할 경우에는 일반적으로 통계청에서 사업체의 규모 층을 구분하는 기준으로 적용하는 방안을 검토하겠다.
- 2009년 환경보호지출계정 조사에서 사용한 산업분류코드와 제9차 산업분류코드의 연계표를 정리하여 아래 <표 6>에 요약하였다. 표본설계 개선안에서 적용할 산업분류는 제9차 분류코드이며 이를 기준으로 조사모집단을 분석할

것이다.

<표 6> 제8차 산업분류와 제9차 산업분류의 연계표

구분	업종	세부업종	기존업종코드	신규업종코드
제조업	음식료(담배제조포함)		15	10-12
	섬유의복모피		17-19	13-15
	목재나무		20	16
	종이인쇄출판	펄프종이	21	17
		출판인쇄	22	18,58(발행업)
	석유정제화학고무	코크스석유정제	23	19
		화합물화학	24	20,21
		고무플라스틱	25	22
	비금속광물		26	23
	제1차금속		27	24
	조립금속기계장비	조립금속	28	25
		기계장비	29-33	26-29
	운수장비		34-35	30-31
가구기타		36-37	32-33	
비제조업	축산재배임업		1-2	1-2
	어업		5	3
	광업		10-12	5-8
	전기가스수도		40-41	35-36
	하수 환경 처리			37-39
	건설업		45-46	41-42
	도소매업		50-52	45-47
	음식숙박		55	55, 56
	운수창고통신	운수창고	60-63	49-52
		통신	64	61
	금융부동산임대		65-67,70-71	64-66,68-69
	보건		85	86
	기타서비스업		72-93	58-60,62,63,70-72,73-75,84-85,87,90-91,94-96

- 담배제조업은 음식료업에 통합하여 분석한다.
- 하수·환경 처리업은 환경전문업체에 해당되며 한국은행의 기업경영분석자료 및 환경산업통계조사 결과를 가공하여 이용하므로 본 기업부문조사에서 추가분석 검토가 필요하다.
- 발행업은 신문발행업(58121), 잡지 및 정기간행물발행업(58122), 정기광고간행물(58123)으로 구성되며 비제조업으로 구분하여 모집단을 분석하고 표본개편 후에 시계열 유지방안 연구에서는 별도 추가 분석이 필요하다.

4.2 일관성 확보 방안

- 2009년 환경보호지출계정 조사에서 사용한 업종구분을 최대한 반영하여 제9

차 한국표준산업분류 코드를 기준으로 업종을 분류하여 신규 표본설계 간의 업종구분 불일치로 인한 시계열 유지에서 발생할 수 있는 문제점을 예방한다.

- 2009년 환경보호지출계정 조사에서는 전수조사층의 구분을 환경오염산업층과 기타로 구분하여 환경오염산업에 관련된 업종의 조사결과의 정확성을 제고하고자 하였으므로 9차 산업분류에서도 반영해야 할 것이다.
- 기존 표본설계에서 업종별로 규모층을 구분한 기준을 알 수 없으므로 종사자 수를 기준으로 규모층 구분방법을 연구하여 관련분야의 전문가 자문을 통해서 결정하고 이로 인한 시계열분석과 일치성 문제가 발생하지 않도록 데이터 시뮬레이션 분석을 통해서 보정하는 방법을 연구한다.

제 5 장 표본추출틀 변경

- 현행 조사에서 사용하고 있는 표본추출틀은 두 종류로 구분된다. 먼저 제조업 대상의 조사는 '03년 기준 「광업·제조업조사」 결과를 이용하고 있고, 비제조업 대상의 조사는 '03년 기준 국세청 법인세 신고업체들을 대상으로 한다. 현행 조사에서 표본추출틀과 관련한 문제점은 이미 2010년 정기통계품질진단에서 지적한 바와 같이 다음 세 가지 사항으로 요약할 수 있다.

- 제조업 및 비제조업간의 조사단위 불일치 문제 해소 필요

그 동안 제조업은 사업체 단위로, 비제조업은 기업체 단위로 조사함으로써 조사단위가 불일치하여 다양한 업종의 사업체로 구성된 기업체에서 환경보호지출계정을 정확하게 파악하기가 곤란하였으므로 이의 개선이 필요하다.

- 환경보호지출계정 표본의 최신화 필요

현행 조사에서 사용하고 있는 제조업의 표본추출틀은 '03년 기준 「광업·제조업조사」 결과이고, 비제조업의 경우는 '03년 기준 국세청 법인세 신고업체들이다. 현행 조사의 표본은 '03년 기준으로 작성된 사업체 리스트로부터 추출되었다. 따라서 현행 조사의 표본사업체는 '03년 이후의 신설 사업체 정보가 반영되지 않은 노후화된 표본이고, '08년에 개정된 산업분류 체계를 제대로 반영하고 있다고 보기 어렵다. 따라서 「환경보호지출계정」 조사를 위한 새로운 표본설계는 통계의 정확도를 좌우할 수 있는 중요하고도 시급한 일이다.

- 환경보호지출계정 조사를 위한 조사대상에 대한 명확한 정의

현행 조사에서 사용하고 있는 제조업의 표본추출틀은 '03년 기준 「광업·제조업조사」 결과이고, 비제조업의 경우는 국세청 법인세 신고업체들이다. 하지만 현행 표본설계 보고서에는 조사대상 업종과 사업체 규모에 대한 명확한 정의와 기준을 알 수 없는 형편이다. 특히 2008년도에 산업분류 체계가 개편됨에 따라 이를 반영한 조사대상 업종과 규모에 대한 명확한 정의와 기준은 이 조사의 기본틀을 구성하는 대단히 중요한 요소이다. 아울러 조사대상의 사업체 규모 범위를 결정할 때는 절사추출법(cut-off sampling)의 관점에서 조사대상을 설정해야 할 것이다. 이와 관련한 참고문헌으로는 Elisson and Elvers (2001), Bee, Benedetti and Espa (2007), 박인호 (2007) 등이 있다.

- 2011년도에 처음 실시된 「경제총조사」는 국내 산업 전체의 생산·고용·비

용 등을 파악하고 이를 통해 적합한 경제정책을 수립하기 위해 전국의 사업체 전체를 동일시점에 통일된 기준으로 조사하는 경제분야 총조사를 말한다. 인구주택총조사(인구센서스)에 비유해 경제센서스라고도 불린다. 조사원이 대상 사업체를 직접 방문해 조사표를 작성하는 면접조사방식으로 진행되나, 사업체에서 원할 경우에는 인터넷조사 및 응답자 직접 기입방식 등을 병행한다. 「경제총조사」는 15종 조사표로 구성되었으며 사업종류, 매출액, 종사자 수, 연간급여액 등 45가지 항목에 대해 조사한다. 우리나라의 첫 경제총조사는 2011년 5월부터 1개월 동안 전국 335.5만 개 사업체 전체를 대상으로 진행됐다. 「경제총조사」는 통계청이 5년마다 실시해 온 「서비스업총조사」와 「산업총조사」를 통합한 것으로, 전수조사가 없었던 농림어업·운수업·건설업 등도 새롭게 포함됐다. 「경제총조사」 결과는 본 연구가 종료될 때까지 공표되지 않은 상황이기 때문에 표본추출틀로 활용할 수 없었다.

- 본 연구에서 사용한 표본추출틀은 「전국사업체조사」 결과이다. 「전국사업체조사」는 한 가지 조사표로 작성하며, 사업체명, 소재지 등 사업체의 기본적인 사항과 조직형태, 사업체구분, 창설연월, 사업의 종류, 종사자 수 등을 조사한다. 따라서 「경제총조사」 결과를 표본추출틀로 활용할 수 있다면 매출액 정보를 표본설계에 반영할 수 있게 되어 효율적인 표본설계가 가능할 것이다.

5.1 「전국사업체조사」 개요

- 「환경보호지출계정」의 기업부문 조사를 위한 표본설계에서 표본추출틀은 2010년 12월 말 기준으로 작성된 「전국사업체조사」 결과이다. 여기서는 「환경보호지출계정」의 기업부문 조사의 조사대상 산업과 사업체 규모 결정을 위한 기초자료를 마련하고자 「전국사업체조사」 결과를 분석한다.
- 본 연구에서 분석 대상은 「전국사업체조사」 결과 중 근로자 1인 이상을 고용하고 있는 전 사업체(국가, 지방행정기관, 군·경찰 및 국·공립교육기관, 가사서비스업(S), 국제기구 및 기타 외국기관(T)은 제외)이다.
- 향후 「전국사업체조사」 자료는 표본설계를 위한 표본추출틀로 활용될 뿐만 아니라 추정단계에서 가중치 계산 등을 위한 기초 자료로 활용된다.

(1) 종사자의 정의

- 「전국사업체조사」에서 종사자는 조사시점인 2010년 12월 31일을 기준하여

해당 사업체에서 업무에 종사한 자를 말한다. 단, 병가자, 단기휴가자, 파업 중인 자, 외국인 취업자, 공공근로요원, 종교단체봉사자 중 상시적으로 봉사하는 자 등은 조사대상에 포함한다. 한편, 조사하고자 하는 해당 사업체에서 급여가 지급되는 종사자를 대상으로 한다. 주거래업체, 지사점, 영업소, 공장 등 산하 다른 사업체로 파견 나간 자, 용역 받은 업무를 위하여 용역을 준 업체에 나가서 일하는 자 등은 종사자에 포함된다. 종사자에 포함되지 않는 경우의 예는 다음과 같다.

- ① 3개월 이상 장기 결근자, 근무 중 군입대자, 전투경찰, 의무경찰, 공익근무요원 등,
- ② 봉사단체의 1일 봉사요원, 각종 단체의 회원, 종교단체의 신자 등,
- ③ 무급 비상근임원,
- ④ 산하 또는 소속기관의 종사자,
- ⑤ 본사·점의 경우 산하 공장, 지사 점의 종사자,
- ⑥ 시·도, 시·군·구의 경우 산하 시·군·구, 읍·면·동 종사자 등,
- ⑦ 조사 해당 사업체에서 급여가 지급되지 않고, 다른 사업체에서 지급되는 자(협력업체, 본사, 본점 등 다른 사업체에서 파견 나온 자, 인력공급업체에서 나온 청소부, 경비원 등 용역업무수행을 위해 용역업체에서 파견 나와 일하는 자 등은 제외)

(2) 종사상 지위에 따른 구분

○ 앞서 정의한 종사자에 대해서 종사상 지위에 따라서 다음과 같이 구분한다.

- ① 자영업주
 - 개인사업체를 소유하면 자신의 책임 하에 직접 경영하는 자
(법인사업체의 대표자는 자영업자가 아님)
- ② 무급가족종사자
 - 무급가족종사자란 자영업주의 가족이나 친인척으로서 임금, 급여를 받지 않고 사업체의 업무를 돕는 자로 정규 근무시간의 1/3이상을 근무한 자
- ③ 상용근로자
 - 고용계약기간이 1년 이상인 근로자 또는 고용계약기간이 정해지지 않고 정규직원으로 일하는 자로서 사업체에서 급여를 지급하는 자
 - 사업체의 유급임원(사장, 대표이사, 전무, 상무, 상근감사 등)
- ④ 임시 및 일일 종사자
 - 고용계약기간이 1년 미만인 근로자로서 사업체에서 급여를 지급하는 자
 - 고용계약은 명시적(explicit) 또는 암묵적(implicit) 계약을 모두 포함함
 - 같은 사업체에 1년 이상 근무하였으나 계약기간이 1개월 이상 1년 미만이거나 처음부터 임시직으로 고용계약이 체결된 사람 또는 매일매일

고용되어 근로의 대가로 일당제 급여를 받고 일하는 사람

⑤ 기타종사자

- 기타 임금근로자, 자영업자 또는 무급가족 종사자로 분류되지 않는 사람 (독자적인 사무실, 점포 또는 작업장이 없고 계약된 사업주에게 종속되어 있지만 스스로 고객을 찾거나 맞이하여 상품이나 서비스를 스스로 직접 제공하고 일한 만큼 실적에 따라 소득을 얻음. 보통 근로제공방법, 근로시간 등은 본인이 독자적으로 결정하는 형태로 일하는 경우 등)

- 「환경보호지출계정」 기업부문 조사의 가중 값 작성단계에서는 조사 기준 시점에서 조사된 근로자(상용근로자+ 임시 및 일일 종사자) 수를 기준하여 사후총화 조정을 적용하는 방안을 고려할 수 있다.

5.2 「전국사업체조사」 결과 분석

(1) 산업대분류별 사업체 및 근로자 수 현황

- <표 7>은 2010년 12월 말을 기준으로 작성된 「전국사업체조사」 결과를 분석한 것이다. <표 7>에서 사업체 규모의 구분은 근로자(상용+ 임시 및 일일) 수에 기초한 것이다. 전체적인 사업체 현황을 보면 근로자(상용+ 임시 및 일일 종사자) 1인 이상을 고용하고 있는 전체 사업체 수는 1,513,540개소로 나타났다. 이들 사업체 중에서 근로자 1~4인이 종사하는 사업체가 1,069,083개소로 전체의 70.6%를 차지하고 있고, 근로자 500인 이상을 고용하고 있는 사업체는 전체의 0.1%인 1,289개소이다.
- 한편 산업대분류별 사업체 현황을 보면 산업대분류에 따라 큰 편차가 있는 것을 알 수 있다. ‘G. 도매 및 소매업’과 ‘I. 숙박 및 음식점업’의 사업체 수가 각각 337,367개소(22.3%)와 287,592개소(19.0%)로 가장 높은 비중을 차지하고 있다.
- <표 8>은 산업대분류 및 사업체 규모별 근로자(상용+ 임시 및 일일) 수 현황이다. 사업체에 종사하고 있는 전체적인 근로자 수는 12,567,146명이다. 이들 중에서 근로자 1~4인 규모의 사업체에 종사하는 근로자는 2,028,078명으로 16.1%에 해당한다. 반면 500인 이상 규모의 사업체에 종사하는 근로자 수는 1,624,094명으로 전체 근로자의 12.9%를 차지하고 있다.
- 한편 산업대분류별 근로자 수 현황을 보면 산업대분류에 따라 큰 편차는 사업체 수에 비해서 작게 나타났다. ‘C. 제조업’과 ‘G. 도매 및 소매업’에 종사하는

근로자 수가 각각 3,075,675명(24.5%)와 1,412,162명(11.2%)로 가장 높은 비중을 차지하고 있다.

(2) 업종 구분별 사업체 및 근로자 수 현황

- <표 9>과 <표 10>은 2010년 12월 말을 기준으로 작성된 「전국사업체조사」 결과를 분석한 것으로 각각 업종 구분 및 사업체 규모별 사업체 수와 근로자 수 현황이다. 여기서 업종 구분은 「환경보호지출계정」 조사의 층화 단계에서 사용된 업종 구분을 제9차 산업분류 체계에 따라 분류하여 사용한 것이다. 제9차 산업분류체계의 개편을 반영하여 ‘담배제조업’과 ‘하수, 폐기물처리, 원료재생 및 환경복원업’을 독립된 업종으로 분리하였다.

(3) 새로운 표본설계에 대한 시사점

- 제9차 산업분류 체계가 2008년부터 적용됨에 따라 층화 단계에서 업종 구분에 대한 재검토가 요구된다. 특히 ‘담배제조업’과 ‘하수, 폐기물처리, 원료재생 및 환경복원업’을 독립된 업종 층으로 구분할 것인가에 대한 검토가 필요할 것이다. 현행 조사에서는 환경부에서 환경 전문 업체를 대상으로 하는 「환경산업통계」의 조사결과를 가공하여 활용하고 있으나 환경보호지출계정에서 필요한 일부 조사항목에 대한 정보 수집을 위해서 민간기업조사에서 ‘하수, 폐기물처리, 원료재생 및 환경복원업’업종을 조사할 경우에는 두 조사의 중복조사 결과의 해석에 유의해야 할 것이다. 아울러 조사대상 업종의 범위도 결정해야 할 것이다. 담배제조업은 전체 사업체수가 13개이므로 별도의 조사업종으로 구분하는 것 보다는 음식료업에 포함하여 층을 구분하는 것이 타당할 것으로 판단하였다. 참고로 ‘하수, 폐기물처리, 원료재생 및 환경복원업’에 대한 사업체 및 근로자 수 현황은 다음과 같다.

① 환경 관련 전문업체 현황

구분	1~4인	5~9인	10~29인	30~99인	100~299인	300~499인	합계
37.하수, 폐수및분뇨처리업	574	254	271	101	10	0	1,210
38.폐기물수집운반, 처리및원료재생업	1,244	923	1,125	398	26	1	3,717
39.환경정화및복원업	20	14	21	7	0	0	62
합계	1,838	1,191	1,417	506	36	1	4,989

② 환경 관련 전문업체에 종사하는 근로자 수 현황

구분	1~4인	5~9인	10~29인	30~99인	100~299인	300~499인	합계
37.하수,폐수및분뇨처리업	1,271	1,689	4,611	4,789	1,535	0	13,895
38.폐기물수집운반, 처리및원료재생업	3,054	6,189	19,000	19,058	4,186	410	51,897
39.환경정화및복원업	56	102	334	420	0	0	912
합계	4,381	7,980	23,945	24,267	5,721	410	66,704

- ‘하수, 폐기물처리, 원료재생 및 환경복원업’은 ‘환경전문업체’의 범주에 속하므로 본조사의 조사모집단에 포함할 경우에는 중복될 수 있으나 기존 「환경산업통계조사」의 제약과 한계(① 자료입수시기, ②조사항목의 불일치)가 있으므로 조사모집단에 포함할 경우에는 중복조사결과에 대한 해석에 유의해야 한다.
- 모집단 분석 결과를 보면 근로자 수 기준으로 살펴볼 때 5인 미만의 규모에 약 71%의 사업체가 집중되어 있는 비대칭적인 형태를 보이고 있다. 근로자 수를 보면 사업체 수의 측면에서 보면 전체 사업체의 약 0.1%인 1,289개소를 차지하고 있는 500인 이상 규모에서 전체 근로자의 12.9%를 차지하고 있다. 환경보호지출 관련 조사항목도 비슷한 경향을 보일 것으로 예상되며 이에 따라 「환경보호지출계정」 조사에서 대규모 사업체가 모총계 추정에서 차지하는 비중이 대단히 높을 것이다. 따라서 대규모 사업체를 대상으로 하는 전수층을 마련하는 방안이 요구된다.
- 현행 「환경보호지출계정」 조사의 표본크기는 제조업 분야에서 1,746개소, 비제조업 분야에서 1,224개소로 전체적으로 약 3,000개소이다. 제한된 표본크기를 효율적으로 활용하여 추정의 정확도를 높이기 위해서는 절사표본조사의 도입을 검토해야 할 것이다. 절사표본조사란 모집단의 일부를 표본조사에서 제외시키는 조사방법이다. 조사변수의 분포가 비대칭적 성격이 강한 사업체조사에서 절사표본추출법이 적용되는 이유는 다음과 같다 (박인호, 2007).
 첫째, 소수의 대규모 사업체들만을 표본조사에 포함하고도 모총합 기준으로 전체 사업체들을 상당 부분 대변할 수 있다는 논리 때문이다.
 둘째, 휴·폐업 및 전업 등과 같은 변동성이 많은 다수의 소규모 사업체들을 조사에서 제외함으로써 표본틀 관리를 용이하게 할 수 있고, 또한 이들의 높은 무응답률로 인한 조사의 어려움을 피할 수 있는 편리성 때문이다.

<표 7> 산업대분류별 사업체 규모별 모집단 사업체 수 현황

구분	1~4인	5~9인	10~29인	30~99인	100~299인	300~499인	500인 이상	합계
A. 농림어업	980	540	529	162	18	0	1	2,230
B. 광업	666	174	287	77	4	2	3	1,213
C1. 경공업	42,866	14,729	11,315	3,514	659	65	31	73,179
C2. 화학공업	11,646	6,863	7,066	2,558	547	51	50	28,781
C3. 금속, 기계	36,430	16,201	16,018	6,375	1,281	114	105	76,524
C4. 전기, 전자, 반도체, 통신	11,687	6,529	6,197	2,466	655	76	108	27,718
D. 전기, 가스, 증기 및 수도사업	497	217	223	402	105	17	14	1,475
E. 하수, 폐기물처리, 원료재생 및 환경복원업	1,838	1,191	1,417	506	36	1	0	4,989
F. 건설업	28,248	16,027	15,850	5,683	1,070	186	153	67,217
G. 도매 및 소매업	270,517	42,695	19,946	3,377	701	93	38	337,367
H. 운수업	18,418	5,384	4,831	2,503	1,139	95	57	32,427
I. 숙박 및 음식점업	253,928	22,875	8,991	1,634	128	21	15	287,592
J. 출판, 영상, 방송통신 및 정보서비스업	10,037	5,234	4,778	2,190	596	90	67	22,992
K. 금융 및 보험업	14,233	9,029	10,491	1,451	192	48	63	35,507
L. 부동산업 및 임대업	32,353	10,969	6,271	893	110	15	9	50,620
M. 전문, 과학 및 기술서비스업	32,213	11,684	7,638	2,568	654	121	108	54,986
N. 사업시설관리 및 사업지원서비스업	14,594	3,816	4,239	2,636	1,084	219	213	26,801
P. 교육서비스업	60,262	11,813	10,020	8,344	1,007	87	131	91,664
Q. 보건업 및 사회복지서비스업	65,307	21,511	12,625	3,659	745	65	115	104,027
R. 예술, 스포츠 및 여가관련 서비스업	31,528	2,534	1,733	781	166	23	7	36,772
S. 협회 및 단체, 수리 및 기타 개인서비스업	130,835	10,444	6,622	1,402	144	11	1	149,459
합계	1,069,083	220,459	157,087	53,181	11,041	1,400	1,289	1,513,540

<표 8> 산업대분류별 사업체 규모별 모집단 근로자 수 현황

구분	1~4인	5~9인	10~29인	30~99인	100~299인	300~499인	500인 이상	합계
A. 농림어업	2,134	3,616	8,267	7,747	2,619	0	774	25,157
B. 광업	1,168	1,184	4,883	3,384	568	752	2,450	14,389
C1. 경공업	89,329	97,628	181,504	172,177	103,001	24,154	21,527	689,320
C2. 화학공업	26,805	46,487	116,515	127,378	86,221	19,631	50,201	473,238
C3. 금속, 기계	79,516	108,898	262,754	317,864	199,490	43,035	236,451	1,248,008
C4. 전기, 전자, 반도체, 통신	26,798	43,974	102,239	123,875	104,689	27,839	235,695	665,109
D. 전기, 가스, 증기 및 수도사업	1,204	1,466	4,435	21,896	16,649	6,090	14,404	66,144
E. 하수, 폐기물처리, 원료재생 및 환경복원업	4,381	7,980	23,945	24,267	5,721	410	0	66,704
F. 건설업	63,537	107,091	258,568	276,801	170,964	70,153	149,092	1,096,206
G. 도매 및 소매업	500,069	271,455	301,061	163,435	109,256	34,740	32,146	1,412,162
H. 운수업	38,885	35,413	79,607	133,444	185,183	35,983	65,866	574,381
I. 숙박 및 음식점업	434,226	145,211	131,929	77,997	21,461	8,778	14,506	834,108
J. 출판, 영상, 방송통신 및 정보서비스업	23,877	34,048	78,103	112,475	94,846	33,382	72,651	449,382
K. 금융 및 보험업	32,886	63,007	152,091	69,597	28,837	18,741	77,558	442,717
L. 부동산업 및 임대업	64,039	71,615	94,269	40,907	16,427	5,314	9,197	301,768
M. 전문, 과학 및 기술서비스업	74,602	74,602	121,111	130,398	108,620	46,383	132,925	688,641
N. 사업시설관리 및 사업지원서비스업	30,143	24,851	70,744	138,301	181,926	84,083	199,085	729,133
P. 교육서비스업	116,616	76,355	166,320	505,067	129,690	33,885	145,421	1,173,354
Q. 보건업 및 사회복지서비스업	167,977	139,849	189,242	190,440	112,526	25,122	153,606	978,762
R. 예술, 스포츠 및 여가관련 서비스업	50,720	16,398	28,441	39,137	24,547	8,400	10,016	177,659
S. 협회 및 단체, 수리 및 기타 개인서비스업	199,166	67,905	101,639	66,406	21,087	4,078	523	460,804
합계	2,028,078	1,439,033	2,477,667	2,742,993	1,724,328	530,953	1,624,094	12,567,146

<표 9> 업종 구분별 사업체 규모별 모집단 사업체 수 현황

업종 구분(환경보호지출계정)		1~4인	5~9인	10~29인	30~99인	100~299인	300~499인	500인 이상	합계	
제조업	음식료	7,772	2,988	2,561	1,165	285	26	14	14,811	
	섬유의복모피	15,723	5,981	4,419	1,311	215	23	11	27,683	
	목재나무	1,979	897	635	104	12	1	1	3,629	
	종이인쇄출판	필프종이	1,989	1,123	1,022	348	70	9	2	4,563
		출판인쇄	6,544	1,649	1,264	296	64	6	7	9,830
	석유정제, 화학, 고무	코크스석유정제	61	74	101	33	2	1	4	276
		화합물화학	2,640	1,856	1,741	801	239	24	23	7,324
		고무플라스틱	6,037	3,331	3,434	1,261	223	13	18	14,317
	비금속광물	2,908	1,602	1,790	463	83	13	5	6,864	
	제1차금속	1,902	1,247	1,575	782	159	21	18	5,704	
	조립금속기계장비	조립금속	20,198	7,451	6,313	1,802	240	13	11	36,028
		기계장비	23,747	12,364	12,037	4,481	966	107	129	53,831
	운수장비	2,270	1,668	2,290	1,776	571	49	55	8,679	
가구 및 기타	9,527	2,409	1,645	389	48	6	3	14,027		
비제조업	축산, 재배, 임업	791	480	486	137	11	0	1	1,906	
	어업	189	60	43	25	7	0	0	324	
	광업	666	174	287	77	4	2	3	1,213	
	전기가스수도	497	217	223	402	105	17	14	1,475	
	하수, 환경, 처리	1,838	1,191	1,417	506	36	1	0	4,989	
	건설업	28,248	16,027	15,850	5,683	1,070	186	153	67,217	
	도소매업	270,517	42,695	19,946	3,377	701	93	38	337,367	
	음식숙박	253,928	22,875	8,991	1,634	128	21	15	287,592	
	운수 및 창고	18,418	5,384	4,831	2,503	1,139	95	57	32,427	
	통신	2,496	1,197	620	447	168	26	12	4,966	
	금융부동산임대	46,586	19,998	16,762	2,344	302	63	72	86,127	
	보건	42,466	9,229	4,547	2,063	625	64	115	59,109	
	기타서비스업	299,146	56,292	42,257	18,971	3,568	520	508	421,262	
합계		1,069,083	220,459	157,087	53,181	11,041	1,400	1,289	1,513,540	

<표 10> 업종 구분별 사업체 규모별 모집단 근로자 수 현황

업종 구분(환경보호지출계정)		1~4인	5~9인	10~29인	30~99인	100~299인	300~499인	500인 이상	합계	
제조업	음식료	15,822	19,886	42,245	59,107	45,856	9,515	10,924	203,355	
	섬유의복모피	34,629	39,508	70,363	63,772	32,570	8,522	6,823	256,187	
	목재나무	4,313	6,018	9,670	4,596	1,749	321	510	27,177	
	종이인쇄출판	펄프종이	4,602	7,491	16,907	17,405	10,752	3,659	1,061	61,877
		출판인쇄	13,000	10,824	19,905	14,181	9,429	2,145	4,199	73,683
	석유정제, 화학, 고무	코크스석유정제	165	536	1,651	1,487	329	474	6,612	11,254
		화합물화학	6,378	12,591	28,794	41,679	37,414	9,166	16,593	152,615
		고무플라스틱	13,970	22,312	57,452	62,014	35,442	4,796	19,782	215,768
	비금속광물	6,292	11,048	28,618	22,198	13,036	5,195	7,214	93,601	
	제1차금속	4,436	8,476	26,094	38,644	24,392	7,981	29,578	139,601	
	조립금속기계장비	조립금속	42,739	49,601	101,821	85,971	36,753	5,215	11,501	333,601
		기계장비	53,814	83,320	197,165	221,958	153,353	39,327	257,340	1,006,277
	운수장비	5,325	11,475	39,913	95,166	89,681	18,351	173,727	433,638	
가구 및 기타	18,561	15,938	26,075	17,996	7,818	2,137	2,209	90,734		
비제조업	축산, 재배, 임업	1,779	3,208	7,523	6,263	1,535	0	774	21,082	
	어업	355	408	744	1,484	1,084	0	0	4,075	
	광업	1,168	1,184	4,883	3,384	568	752	2,450	14,389	
	전기가스수도	1,204	1,466	4,435	21,896	16,649	6,090	14,404	66,144	
	하수, 환경, 처리	4,381	7,980	23,945	24,267	5,721	410	0	66,704	
	건설업	63,537	107,091	258,568	276,801	170,964	70,153	149,092	1,096,206	
	도소매업	500,069	271,455	301,061	163,435	109,256	34,740	32,146	1,412,162	
	음식숙박	434,226	145,211	131,929	77,997	21,461	8,778	14,506	834,108	
	운수 및 창고	38,885	35,413	79,607	133,444	185,183	35,983	65,866	574,381	
	통신	6,689	7,299	10,725	24,711	28,457	9,600	14,029	101,510	
	금융부동산임대	96,925	134,622	246,360	110,504	45,264	24,055	86,755	744,485	
	보건	104,767	58,552	68,172	117,624	96,305	24,674	153,606	623,700	
	기타서비스업	550,047	366,120	673,042	1,035,009	543,307	198,914	542,393	3,908,832	
합계		2,028,078	1,439,033	2,477,667	2,742,993	1,724,328	530,953	1,624,094	12,567,146	

5.3 일관성 확보 방안

- 제9차 산업분류 체계가 2008년부터 적용됨에 따라 층화 단계 뿐 만 아니라 통계 작성 단계에서 제8차 산업분류 체계와 제9차 산업분류 체계의 정밀한 연계 작업이 필요하다. 아울러 조사대상 범위에 대한 전문가 논의과정이 요구된다. 대상 산업을 제한하거나 사업체 규모를 제한하여 제한된 표본크기를 효율적으로 활용할 수 있는 방안을 모색해야 할 것이다.
- 앞서 기술한 바와 같이 「환경보호지출계정」을 작성할 때 환경 전문업체를 대상으로 하는 환경부의 「환경산업통계」의 조사결과를 가공하여 활용하고 있다. 이에 따라 원칙적으로 사업체 대상의 조사에서 환경 전문업체는 중복되는 측면이 있으므로 조사대상 포함여부가 검토되어야 한다. 이들 환경 전문업체가 「환경보호지출계정」에 미치는 영향이 대단히 크다는 점을 고려하여 통계작성 방법과 두 조사에서 중복 조사 가능성 등을 심도 있게 검토하여 표본 설계 방안을 마련해야 할 것이다. 환경 전문업체는 앞서 제시한 ‘37. 하수, 폐수 및 분뇨처리업’, ‘38. 폐기물 수집 운반, 처리 및 원료 재생업’, ‘39. 환경정화 및 복원업’ 이외의 산업분류에도 포함될 수 있다는 점에서 조사 중복 가능성을 고려하여 조사대상에 포함했을 경우에는 중복조사에 의한 결과에서 오류가 발생할 수 있으므로 해석에서 유의해야 할 것이다.

제 6 장 최적표본설계방안 : 신규표본설계의 표본추출 가이드라인

6.1 표본설계의 기본원칙

- 그 동안 「환경보호지출계정」의 기업부문 조사에서 제조업은 사업체 단위로, 비제조업은 기업체 단위로 조사함으로써 조사단위가 불일치하여 다양한 업종의 사업체로 구성된 기업체에서 환경보호지출계정을 정확하게 파악하는 것이 곤란하였다. 새로운 표본설계에서는 「환경보호지출계정」의 기업부문 조사에서 전체 산업분류에 대하여 조사단위를 사업체로 통일하여 조사단위의 불일치로 발생할 수 있는 측정오차를 줄이는 것이 바람직할 것이다. 표본추출틀은 좀 더 다양한 사업체 관련 정보를 활용할 수 있는 「경제총조사」 결과를 이용하는 것을 원칙으로 한다. 통계공표의 지연으로 「경제총조사」 결과를 활용할 수 없다면 「전국사업체조사」 결과를 표본추출틀로 활용한다.
- 현행 표본설계 보고서는 조사대상 업종과 사업체 규모에 대한 명확한 정의와 기준에 대한 기술이 없는 형편이다. 특히 2008년도에 산업분류 체계가 개편됨에 따라 이를 반영한 조사대상 업종과 규모에 대한 명확한 정의와 기준은 이 조사의 조사모집단을 결정하는 대단히 중요한 과정이다. 새로운 표본설계는 조사대상 업종과 규모에 대한 명확한 정의와 기준을 마련하여 조사모집단을 명확하게 정의하는 작업으로부터 시작되어야 한다.
- 「환경보호지출계정」 품질진단에서 전문가 표적집단 면접 결과(품질진단보고서 30쪽 참조)에 의하면 해당 분야 통계이용자들은 산업 세부 업종별 등 세세 분류 항목의 자료가 제공되기를 요구하고 있다. 이와 같은 통계이용자들의 요구를 충족시키기 위해서는 업종 구분을 우선적으로 고려한 현행의 층화 방법은 여전히 유효한 층화방안으로 볼 수 있다. 다만, 현재의 업종 구분이 이용자 관점에서 유사한 산업분류로 구성되었는지에 대한 검토가 필요하다.
- 「환경보호지출계정」 결과의 통계이용자들은 산업 세부 업종별 통계를 요구하고 있다. 한편, 앞선 <표 3>에 제시한 2009년 「환경보호지출계정」 조사데이터 분석결과를 보면 환경보호경상지출액이나 인건비의 업종별 편차가 대단히 큰 것을 알 수 있다. 따라서 층화에 사용된 업종 구분은 통계작성단위의 역할을 하게 된다. 따라서 각 업종의 중요도를 고려해 업종별 추정의 목표상 대표준오차를 결정한 후 이를 기초로 업종별 표본크기를 결정하는 것이 바람직할 것이다. 전체 표본크기는 각 업종별 표본크기의 합계로 구한다.

- 「환경보호지출계정」 조사는 그 속성상 환경보호경상지출액이나 인건비의 사업체 규모별 편차가 대단히 커서 주요 조사변수에 대한 왜도가 대단히 큰 값을 갖게 된다. 이와 같은 경우에 현행 표본설계와 마찬가지로 전수조사층과 표본조사층을 구성하여 표본설계하는 것은 추정의 정확도를 높일 수 있는 효과적인 방안이다.

6.2 변형절사추출법의 적용 방안

- Lavallée & Hidiroglou (1988)은 왜도가 대단히 큰 모집단(오른쪽으로 꼬리가 긴 분포)에 대해서 하나의 전수조사층(take-all stratum)과 몇 개의 표본조사층(take-some stratum)을 구성하고자 할 때 층 경계점을 구할 수 있는 알고리즘을 제시하였다. Lavallée & Hidiroglou 방법(절사층화추출법)은 추정에 대한 목표상대표준오차와 표본조사층에 대한 표본배분법(예: 멱배분이나 네이만배분 등)이 정해진 상황에서 각 표본조사층에서 단순임의추출법을 적용할 때 목표오차를 만족하면서 전체 표본크기를 최소화할 수 있는 층 경계점을 결정하는 것을 목적으로 한다. 이 방법에 의한 층 경계점 결정방법은 Hidiroglou(1986)과 Sethi(1963)이 제안했던 층 경계점 결정방법을 조합한 형태로 볼 수 있다. Lavallée & Hidiroglou 방법은 전수조사층이 있는 층화임의추출법에서 표본크기 결정 공식을 최소로 하는 층 경계점인 $b_{(1)}, b_{(2)}, \dots, b_{(L-1)}$ 을 반복계산법(iterative method)에 따라 구할 수 있는 알고리즘이다. 다음 식에서 W_h, S_h^2, \bar{Y}_h 등은 모두 층 경계점 $b_{(1)}, b_{(2)}, \dots, b_{(L-1)}$ 의 함수로 볼 수 있다.

$$n = N \cdot W_L + \frac{N \left(\sum_{h=1}^{L-1} W_h^2 S_h^2 / a_h \right)}{N(c v_0 \bar{Y})^2 + \sum_{h=1}^{L-1} W_h S_h^2}$$

여기서, L 은 전체 층의 수, $c v_0$ 는 목표상대표준오차, $a_h = \frac{N_h^p}{\sum_{h=1}^{L-1} N_h^p}$, ($0 < p < \infty$)

이다.

- Rivest (2002)는 Lavallée & Hidiroglou 방법을 일반화하여 전수조사층과 표본조사층의 경계점의 결정을 위한 알고리즘을 제안하였다. 이 방법은 표본설계 단계에서 고려한 층화변수와 주관심 변수가 대개 서로 관련성은 높지만 일치하지는 않기 때문에 나타날 수 있는 주관심 변수 추정량의 목표상대표준오차(표본설계 단계에서 설정)와 실제 상대표준오차의 차이를 줄이기 위한 방안으로 제시되었다. 이 방안은 층화변수와 주관심 변수 간의 로그변환된 선형

모형의 모형화 방법을 통해서 Lavallée & Hidiroglou 방법을 일반화하는 층 경계점 결정을 위한 알고리즘이다.

- 새로운 표본설계에서 표본조사층 및 전수조사층의 층 경계점은 앞서 제시한 Lavallée & Hidiroglou 방법을 적용하는 것을 검토해야 할 것이다.

6.3 층화 변수에 대한 검토

- 표본추출틀로 「경제총조사」 결과를 이용할 수 있다면 층화변수로 매출액 정보를 활용하여 Lavallée & Hidiroglou 방법을 적용하여 전수층과 표본층의 층 경계점을 구할 수 있다. 하지만 공표일정의 지연으로 「경제총조사」 결과를 표본추출틀로 활용할 수 없는 경우에는 「전국사업체조사」 결과를 이용하여 표본설계를 진행해야 한다. 「전국사업체조사」 결과에는 매출액 정보를 포함하고 있지 않기 때문에 사업체 규모를 나타내는 보조변수인 근로자 수 정보를 층화변수로 이용하여 Lavallée & Hidiroglou 방법을 적용해야 할 것이다. 다른 방안으로 Rivest (2002)이 제안한 일반화 Lavallée & Hidiroglou 방법을 적용하는 것도 고려해야 할 것이다. 이 방안은 매출액과 사업체 특성변수(근로자 수)의 로그선형모형을 이용하여 Lavallée & Hidiroglou 방법을 개선한 것이다.
- 일반화 Lavallée & Hidiroglou 방법을 적용하기 위해서는 현재 조사표에 대한 개선이 필요하다. 우선 추가 조사항목으로 표본사업체에 대한 산업분류와 근로자 수 정보가 조사표에 포함되어야 한다. 두 조사항목이 추가로 조사되면 매출액과 사업체 내 근로자 총수 간의 로그선형모형을 통해서 일반화 Lavallée & Hidiroglou 방법의 적용이 가능하게 된다.

6.4 표본크기 결정의 원칙

- 「환경보호지출계정」 조사에서 층화에 사용된 업종 구분은 통계작성단위의 역할을 하게 된다. 따라서 각 업종의 중요도를 고려해 업종별 추정의 목표상대표준오차를 결정한 후 이를 기초로 업종별 표본크기를 결정하도록 한다. 전체 표본크기는 각 업종별 표본크기의 합계로 구한다.
- 각 업종 층에서 표본크기는 목표상대표준오차가 주어진 상황에서 Rivest (2002)의 일반화 Lavallée & Hidiroglou 방법을 적용하여 결정하는 것을 원칙으로 한다. 매출액과 사업체 내 근로자 총수 간의 로그선형모형을 적합할 수 없는 경우에는 근로자 수를 층화변수로 하는 Lavallée & Hidiroglou 방법

을 적용하여 결정한다.

- 2009년 「환경보호지출계정」 조사데이터에서 계산한 업종별 모평균추정량의 변동계수는 <표 3>에 주어졌으며 제조업과 비제조업간의 차이가 있을 볼 수 있다. 예를 들어 인건비 항목에 대한 제조업종의 변동계수 평균은 18.90%이고 비제조업의 변동계수 평균은 34.84%이며, 환경보호경상지출 항목에서는 제조업의 변동계수 평균은 25.7%이고 비제조업의 변동계수 평균은 38.45%로 인건비에서 변동계수가 더 작음을 보이고 있다. 이런 조사모집단의 특성을 반영하여 제조업의 업종별 목표변동계수는 15%로하고 비제조업의 목표변동계수는 30%로 설정하여 필요한 표본크기를 아래 식으로 계산하였다.

$$n_{2012j} = n_{2009j} \cdot \left(\frac{CV_{2009j}}{CV_{2012j}} \right)^2, \text{ 여기서 } j \text{는 업종을 나타냄}$$

위 식으로 산출한 업종별 표본크기가 기존 표본크기에 비해서 대폭 축소된 업종에 대해서는 신규표본에서 표본크기의 축소 폭이 클 경우에 모수추정량에서 시계열유지에 문제가 있을 수 있으므로 기존 표본크기와 신규표본크기의 평균으로 보정하였고 마지막으로 업종별 최소표본크기를 30이상으로 보정한 업종별로 필요한 표본크기를 결정한 결과를 <표 11>에 주어졌으며 전체 표본크기는 4,120개이며 2009년 응답한 표본사업체 2,519개보다는 63.5%가 증가하였다. 위에서 설명한 표본크기 결정절차를 아래와 같이 3단계로 정리할 수 있다.

- ① 업종별 표본크기 산출 : 위에 주어진 표본크기 산출식에서 업종별로 목표 변동계수를 설정하여 필요한 표본크기를 계산한다.
- ② 업종별로 신규로 계산된 표본크기와 2009년도 조사된 표본크기를 비교하여 신규 표본크기가 매우 작을 경우에는 평균값 수준으로 보정한다.
- ③ 업종별로 최소 표본크기가 30이상이 될 수 있도록 조정한다.

<표 11>에는 2010년 전국사업체조사를 추출틀로 했을 경우에 업종별로 산출한 모집단 크기와 2009년에 조사에 응답한 데이터의 표본크기도 함께 주어졌으며 ‘인건비’와 ‘환경보호경상지출’에 대한 업종별로 목표변동계수(제조업 15%, 비제조업 30%)를 기준으로 계산한 표본크기도 함께 주어졌다. ‘인건비’와 ‘환경보호경상지출’에 대해서 다양한 크기의 목표변동계수들을 기준으로 필요한 표본크기를 업종별로 산출한 결과를 <부록 1>에 참고할 수 있도록 포함하였다.

<표 11> 표본개선안의 조사모집단과 필요 표본크기와 기대변동계수

업종분류		모집단 크기	표본크기		목표CV(제조15,비제조30)		기대변동계수	
			개선안	기존	인건비	EPEA	인건비	EPEA
음식류		14811	94	141	44	74	10.19	13.30
섬유의복모피		27,683	97	104	97	203	15.00	21.65
목재나무		3,629	69	43	69	235	14.90	27.39
종이인쇄출판	펄프종이	4,563	54	70	44	66	13.54	16.48
	출판인쇄	9,830	310	97	310	163	14.76	10.70
석유정제화학고무	코크스석유정제	276	40	26	159	253	27.70	34.87
	화합물화학	7,324	185	177	185	198	14.81	15.34
	고무플라스틱	14,317	582	151	584	422	14.72	12.50
비금속광물		6,864	109	148	109	323	14.88	25.60
제1차금속		5,704	280	130	280	1923	14.64	38.33
조립금속기계장비	조립금속	36,028	100	139	100	140	14.96	17.70
	기계장비	53,831	281	211	281	1069	15.00	29.26
운수장비		8,679	174	99	174	214	14.86	16.45
가구기타		14,027	158	122	158	232	14.91	18.06
축산재배임업		1,906	51	61	8	41	11.67	26.49
어업		324	30	16	3	8	8.75	14.99
광업		1,213	92	32	158	92	37.76	28.82
전기가스수도		1,475	39	40	38	39	29.39	29.67
건설업		67,217	222	103	116	222	21.67	29.96
도소매업		337,367	310	161	329	310	30.90	30.01
음식숙박		287,592	167	78	119	167	25.30	30.03
운수창고통신	운수창고	32,427	162	64	105	162	24.10	29.93
	통신	4,966	30	11	0	16	0	21.96
금융부동산임대		86,127	196	98	324	196	38.51	29.97
보건		59,109	74	102	18	44	14.76	23.13
기타서비스업		421,262	214	95	53	214	14.96	29.98
합계		1,508,551	4,120	2,519	3,865	7,026		

6.5 특이치 검출 방안

- 「환경보호지출계정」 조사의 조사항목은 그 속성상 특이치가 빈번하게 발생할 가능성이 있다. 따라서 특이치를 검출하여 조사 값을 점검하는 것은 조사의 정확도를 높이는 중요한 과정이다. 특이치 검출 방법은 한 변수를 사용한 검출 방법과 두 변수간의 관계를 이용한 검출 방법으로 구분된다.

(1) range edit 방법

먼저 한 변수에 의해 특이치를 검출하는 방법으로 m 이 표본평균이고 s 가 표

본표준편차일 때

$$d_i = \frac{y_i - m}{s}$$

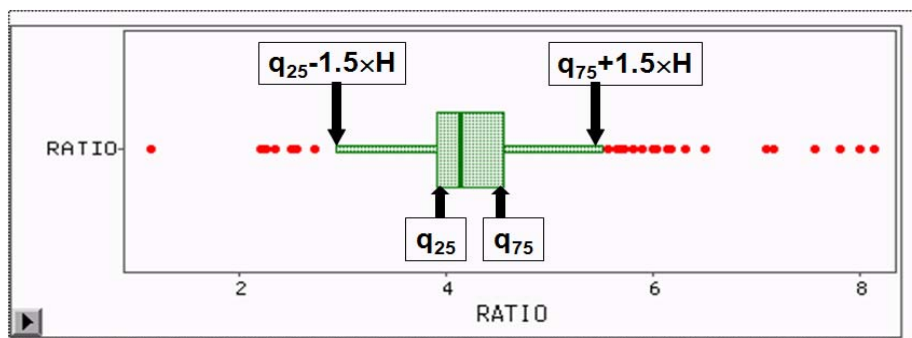
을 사용하여 d_i 가 임의의 C 값에 대하여 $|d_i| > C$ 일 때 특이치로 검출된다. 이 방법은 경미한 특이치를 검출하지 못하는 masking effect 현상이 발생한다. 그래서 robust한 방법으로 표본평균 대신 중앙값을 사용하고 표본표준편차를 대신하여 사분위수 범위나 $MAD = \text{median}|x - \text{median}(x)|$ 을 사용한다. 두 번째로 한 변수에 의해 특이치를 검출하는 방법으로 Quartile 방법인데 q_1 이 조사변수에 대한 제 1사분위수이고 q_2 가 중앙값이며 q_3 가 제 3사분위수일 때

$$[q_2 - C_L(q_3 - q_1), q_2 + C_U(q_3 - q_2)]$$

의 범위를 벗어나면 특이치로 구분하는 것이다. 여기서 C_L 과 C_U 는 과거 조사데이터에 대한 경험적 분석(empirical analysis)을 통하여 결정된다(Bernier and Nobrega, 1998).

(2) resistant fences methods

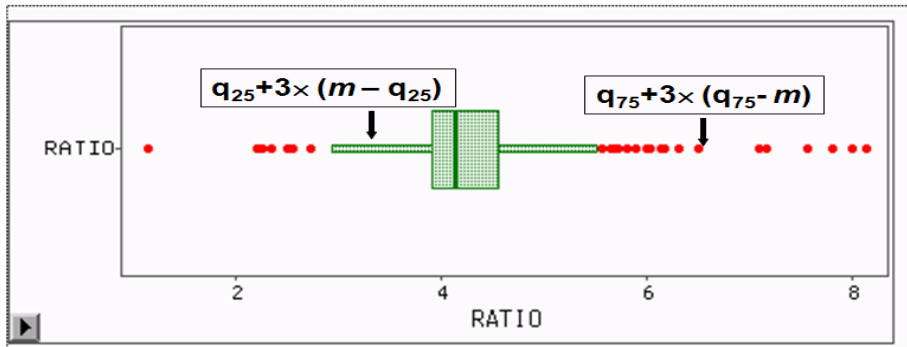
두 변수의 ratio 값에 대한 q_1 가 제 1사분위수고 q_3 가 제 3사분위수일 때 $(q_1 - k \times (q_3 - q_1), q_3 + k \times (q_3 - q_1))$ 의 구간 밖에 값이 존재할 때 그 조사 값은 특이치로 분류한다. 여기서 k 값은 1.5, 2, 3을 사용한다. 이 방법은 모집단의 분포가 대칭적(symmetrical) 분포일 때 사용이 가능하다. [그림 1]은 사업체 대상의 조사 자료에서 annual payroll과 total employment의 두 변수의 비의 값을 가지고 특이치를 찾는 사례이다 (Thompson, 2007).



[그림 1] resistant fences methods (Thompson, 2007)

(3) Asymmetric resistant fences methods

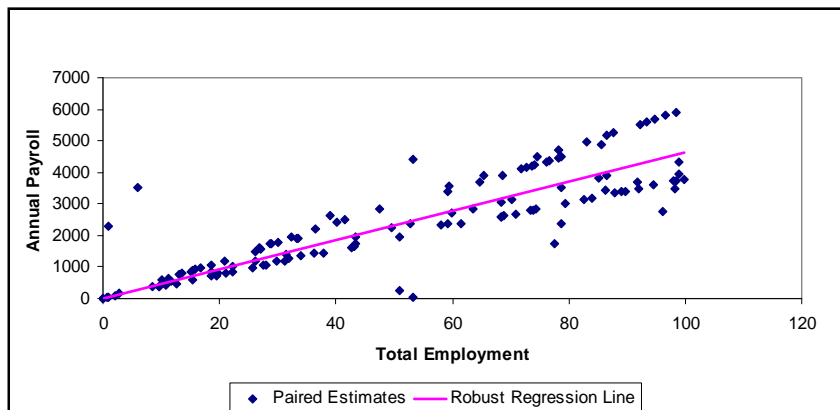
모집단의 분포가 비대칭적(skewed) 분포일 때 적용할 수 있는 방법이다. 두 변수의 ratio 값에 대한 m 값이 중앙값일 때 $(q_1 - k \times (m - q_1), q_3 + k \times (q_3 - m))$ 의 범위 밖에 존재하는 값을 특이치라 규정한다. 여기서 k 의 값은 3, 4, 6을 사용한다. Thompson(2007)에서 제시된 ratio값에 대해 Asymmetric resistant fences methods 방법을 적용한 그래프는 아래와 같다.



[그림 2] Asymmetric resistant fences methods (Thompson, 2007)

(4) Robust regression

최소절사제곱추정량(least trimmed squares estimator : LTS)을 사용하여 least trimmed squares robust regression을 적용한다. 잔차의 값이 $|잔차| \geq 3 \times robust\ MSE$ 이면 특이치로 구분한다(Thompson, 2007).



[그림 3] Robust regression (Thompson, 2007)

(5) Hidiroglou and Berthelot 방법

이의규(2010)는 광업 및 제조업 조사에 Hidiroglou and Berthelot 방법을 사용하여 특이치를 검출하였다. Hidiroglou and Berthelot 방법은 ratio edit 방법의 일종이며 검출 순서는 다음과 같다.

$$\textcircled{1} r_i = \frac{x_{i,t+1}}{x_{i,t}}, \text{ 여기서 } x_{i,t} \text{ 는 } t \text{ 시간에 } i \text{ 번째 값이다.}$$

$$\textcircled{2} s_i = \begin{cases} 1 - \frac{r_{median}}{r_i}, & 0 < r_i < r_{median} \\ \frac{r_i}{r_{median}} - 1, & r_i \geq r_{median} \end{cases} \text{ 을 계산한다. 단 } r_{median} \text{ 은 } r_i \text{ 의 중앙값}$$

이다.

$$\textcircled{3} e_i = s_i \{ \text{MAX}(x_{i,t}, x_{i,t+1}) \}^u \text{ 을 계산한다. 단 } \text{MAX}() \text{ 는 최대값을 계산하는 함수이고, } 0 \leq u \leq 1 \text{ 이며 } u=1 \text{ 이면 큰 단위의 작은 변동에 더 큰 의미를 부여하는 수치이다.}$$

$$\textcircled{4} d_{D_L} = \text{MAX}(e_{median} - e_{q_1}, |ae_{median}|) \text{ 와 } d_{D_U} = \text{MAX}(e_{q_3} - e_{median}, |ae_{median}|) \text{ 를 계산한다. 단 } e_{q_1}, e_{median}, e_{q_3} \text{ 는 각각 1사분위수, 중앙값, 3사분위수이며 } a=0.05 \text{ 가 추천된다.}$$

$$\textcircled{5} \text{ 특이치의 판별범위는 } (e_{median} - c \times d_{D_L}, e_{median} + c \times d_{D_U}) \text{ 의 구간 밖에 위치한 것을 말한다. 일반적으로 } c=6 \text{ 이 사용된다.}$$

Hidiroglou and Berthelot 방법의 확장은 위에 제시된 r_i 대신에

$$r_{curr,i} = \frac{x_{i,t}}{y_{i,t}}$$

을 이용할 수 있으며, e_i 대신에

$$e_i = s_i \{ \text{MAX}(x_{i,t}, r_{curr,m} \times y_{i,t}) \}^u$$

을 사용할 수 있다. 단 $r_{curr,m}$ 은 $r_{curr,i}$ 의 중앙값이다.

- 현행 「환경보호지출계정」 조사에서는 해당년도와 전년도의 조사항목 값을 함께 조사하고 있다. 따라서 중요 조사항목에 대한 특이치 검출은 Hidiroglou and Berthelot 방법을 사용하는 것을 검토해야 할 것이다.

6.6 가중치 작성 및 추정방법

- 사업체 가중값을 구하기 위해서는 설계 가중값과 무응답 조정의 과정을 거쳐야 한다. 우선 설계 가중값과 무응답 조정값을 구하여 표본 사업체에 대한 가

중값을 구하고, 다음 단계로 벤치마킹 과정(사후층화 조정)을 거쳐야 할 것이다. 표본설계에서 사후층화 단계에서 사용되는 벤치마킹 정보는 활용할 수 있는 최신 정보의 「전국사업체조사」나 「경제총조사」 결과의 근로자 수와 매출액 정보를 이용한다.

- 이 조사에서 주요 조사항목에 대한 모총계 및 모평균 추정은 가중치를 이용한 추정량을 이용하여 추정한다. 모평균에 대한 추정량은 표본가중평균으로 아래 식으로 계산한다.

$$\hat{\mu}_h = \frac{\sum_{i=1}^{n_{hi}} \sum_{j=1}^{n_{ij}} w_{hij} y_{hij}}{\sum_{i=1}^{n_{hi}} \sum_{j=1}^{n_{ij}} w_{hij}}$$

여기서, h 는 업종을 나타내고 i 는 규모층을 나타내며 j 는 표본사업체를 의미하고 w_{hij} 와 y_{hij} 는 각각 h 업종 i 규모층 내의 j 사업체의 가중치와 관찰값을 나타낸다.

- 모총계에 대한 추정량은 다음 식으로 계산한다.

$$\hat{\tau}_h = \sum_{i=1}^{n_{hi}} \sum_{j=1}^{n_{ij}} w_{hij} y_{hij}$$

여기서, h 는 업종을 나타내고 i 는 규모층을 나타내며 j 는 표본사업체를 의미하고 w_{hij} 와 y_{hij} 는 각각 h 업종 i 규모층 내의 j 사업체의 가중치와 관찰값을 나타낸다.

- 표본조사 데이터를 분석할 때 가중치를 무시하고 분석하면 모수 추정에 심각한 편향(bias)이 발생할 수 있고, 추정량의 분산이 과소평가되어 문제가 될 수 있다. 각종 총계 추정량에 대한 분산이나 상대표준오차 추정값은 복합표본조사 데이터를 분석할 수 있는 전문 통계패키지인 SAS, SPSS 등을 이용해서 구한다.

6.7 2013년 조사를 위한 표본설계 연구절차

표본조사를 위해 모집단을 잘 대표하는 표본을 뽑고 조사를 통해서 얻어진 결과를 기초로 모수를 추정하는 전체 과정을 표본설계(design of sample survey)라고 한다. 마치 건축설계사가 건물을 설계하듯이 표본조사에서도 모집단을 효과적으로 대표할 수 있는 표본을 추출할 수 있는 설계를 해야 한다. 표본설계는 작성되는 통계의 품질에 직접적으로 영향을 미치며, 표본조사를 통해서 얻게 되는 통

계가 요구하는 정확도, 예산상의 제약, 표본추출 방법의 이용가능성, 활용 가능한 보조정보의 수준, 이용될 조사기법 등 다양한 요소들을 고려하여 이루어지게 된다.

주어진 여건 하에서 가장 경제적이고 정확성이 높으며 효율적인 표본을 설계하는 것이 표본설계의 목표이다. 표본설계 과정에서 중요하게 다루어야 할 중요한 주제로는 조사단위 및 조사대상의 결정, 모집단과 추출틀 결정, 층화 방법 및 경계점, 표본크기 결정 및 표본배분, 표본추출법, 모수 추정식 및 분산 추정방법, 가중치 작성법, 표본관리 등이 있다. 여기서는 「환경보호지출계정」 조사를 위한 표본설계 연구의 전체적인 절차를 요약하여 제시한다.

(1) 조사단위 결정

표본조사에서 필요한 정보를 얻기 위해 조사를 할 때 조사의 대상이 되는 가장 최소의 요소를 조사단위 또는 관찰단위라고 부르기도 한다. 표본조사에서 조사단위를 어떤 것으로 할 것인가는 조사목적과 조사 가능성을 검토하여 결정된다. 「환경보호지출계정」을 위한 조사에서 현행 표본설계의 조사단위는 제조업의 경우 사업장 단위를 사용하고 있고, 비제조업의 경우 기업체 단위를 사용하고 있다. 이와 같이 사업체와 기업체가 조사단위로 혼재된 경우에는 포괄범위가 서로 다름에 따라 측정오차 등의 비표본오차의 발생 가능성이 있고, 특히 다양한 업종의 사업체로 구성된 기업체에서 환경보호지출계정을 정확하게 파악하기가 곤란하다는 점을 고려한 것이다. 따라서 새로운 표본설계에서는 제조업과 비제조업에서 모두 사업체를 조사단위로 하여 표본설계하는 것이 바람직할 것이다. 다만, 표본설계 과정에서는 일부 사업체를 대상으로 하는 사전조사를 통해 현행 조사표의 각 조사항목에 대해서 사업체 수준에서의 정확한 응답이 가능한가와 응답이 곤란한 조사항목은 없는가에 대하여 검토하여 조사항목에 대하여 수정·보완하는 과정이 필요할 것이다.

(2) 모집단 및 표본추출틀 결정

통계조사에서 연구의 대상이 되는 모든 조사단위들의 전체 집합을 목표모집단(target population)이라고 한다. 일반적으로 목표 모집단을 그대로 나타내는 표본추출틀을 작성하기가 현실적으로 대단히 어려운 것이 보통이다. 이럴 때 현실적으로 조사가능하고, 표본추출틀을 작성할 수 있는 추출단위만으로 구성된 모집단을 조사모집단(study population)이라고 한다. 조사가능 모집단은 목표 모집단과 정확하게 일치하지는 않더라도 상당히 유사해야 하며 실제 활용할 수 있다. 표본추출틀(sampling frame)은 모집단에 속하는 모든 추출단위들의 목록이다. 올바른 추출틀이 되려면 정의된 모집단내의 모든 추출단위들이 빠짐없이 다 망

라되어 있어야 하고 또 어느 추출단위도 중복된 것이 없어야 한다.

「환경보호지출계정」 조사에서 목표모집단은 조사시점 현재 대한민국 행정권이 미치는 전 지역에 소재하는 모든 사업체들이다. 현행 표본설계에서는 환경오염방지출 규모가 미미할 것으로 보이는 ‘재배업’, ‘수렵업 및 임업’, ‘도소매업 중 자동차 판매 및 수선업’, ‘금융보험업’, ‘부동산 및 사업서비스업’, ‘사회 및 개인 서비스업 중 보건업’ 등을 제외하여 조사모집단을 구성하였다. 현행 조사의 표본추출틀은 1992년도 국세청의 법인세 신고대상업체 중 법인 기업을 표본추출틀로 하였다. 1992년도 법인세 신고기업 94,883개 중 비조사대상업종 기업, 매출액이 없는 기업 등을 제외한 49,554개 기업을 표본추출틀로 하였다.

현행 조사에서 사용하고 있는 표본추출틀은 두 종류로 구분된다. 먼저 제조업 대상의 조사는 '03년 기준 「광업·제조업조사」 결과를 이용하고 있고, 비제조업 대상의 조사는 '03년 기준 국세청 법인세 신고업체들을 대상으로 하고 있다(통계청 품질진단보고서, 21쪽 참조).

새로운 표본설계는 조사대상 업종과 규모에 대한 명확한 정의와 기준을 마련하여 조사모집단을 명확하게 정의하는 작업으로부터 시작되어야 한다. 특히 2008년도에 산업분류 체계가 개편됨에 따라 이를 반영한 조사대상 업종과 규모에 대한 명확한 정의와 기준은 이 조사의 기본틀을 구성하는 대단히 중요한 요소이고, 향후 이 조사의 시계열 유지 측면에서도 대단히 중요하다.

새로운 표본설계에서 조사대상 업종과 규모(종사자 수 또는 매출액 기준으로 조사대상 사업체 결정)의 결정은 조사 가능성과 효율성의 검토를 통해서 이루어져야 할 것이다. 사업체 규모를 기준으로 조사대상을 결정할 때는 앞서 소개한 절사표본추출법에 대한 검토가 필요하다. 절사표본조사란 모집단을 구성하는 사업체 중 규모가 매우 큰 일부 사업체는 전수조사하고, 규모가 중간인 사업체들은 표본조사를 하며 규모가 매우 작은 일부 사업체들은 조사대상에서 제외하는 조사방법이다 (박인호, 2007). 여기서 전수조사와 조사대상에서 제외하는 기준은 모집단의 특성에 따라 정해지며 세부적인 알고리즘은 Lavallée & Hidiroglou (1998)에 주어졌다.

아울러 현행 조사와 새로운 표본설계에 의한 조사의 시계열 연계를 위해서는 현행 표본설계에 대한 충분한 검토와 2008년도 산업분류 체계 개편의 내용(산업분류 개편에 따른 시계열 유지 방안은 앞서 3.3절에서 다루었음)을 반영할 수 있도록 표본설계 되어야 한다.

새로운 표본설계에서 표본추출틀은 좀 더 다양한 사업체 관련 정보를 활용할 수 있는 「경제총조사」 결과를 이용하는 것을 원칙으로 한다. 신규 표본설계 과정에서 통계공표의 지연으로 「경제총조사」 결과를 활용할 수 없다면 「전국사업체조사」 결과를 표본추출틀로 활용한다. 한편, 환경보호지출계정에서 차지하는 비중이 높은 환경오염관련 산업을 부차모집단으로 간주하여 독립적으로 관리하고 현행의 표본설계하는 방안은 여전히 유효한 방안으로 판단된다. 따라서 'E. 하

수, 폐기물처리, 원료재생 및 환경복원업'은 '환경전문업체'의 범주에 속하므로 본조사의 조사모집단에 포함할 경우에는 중복될 수 있으므로 조사모집단에서 제외하는 것이 바람직하다.

(3) 층화 변수 선택

「환경보호지출계정」 품질진단에서 전문가 표적집단 면접 결과(품질진단보고서 30쪽 참조)에 의하면 해당 분야 통계이용자들은 산업 세부 업종별 등 세세분류 항목의 자료가 제공되기를 요구하고 있다. 이와 같은 통계이용자들의 요구를 충족시키기 위해서는 업종 구분을 우선적으로 고려한 현행의 층화 방법은 여전히 유효한 층화방안으로 볼 수 있다.

현행 조사에서 각 업종 구분별 추정의 정확도를 평가하기 위해서는 주요 관심변수에 대한 추정량의 상대표준오차를 구하여 살펴보아야 한다. 이와 같이 현행 조사데이터를 분석하는 것은 층화, 목표오차 설정, 표본크기 결정, 표본배분법 결정 등의 과정에서 필수적인 정보를 제공해 준다.

새로운 표본설계의 층화 기준은 업종 구분과 사업체 규모(매출액 또는 종사자 수 기준)를 고려하여 층화하는 방안을 검토해야 한다. 다만, 현재의 업종 구분이 이용자 관점에서 유사한 산업분류로 구성되었는지에 대한 검토가 필요하다.

앞선 모집단 분석 결과를 보면 근로자 수 기준으로 살펴볼 때 사업체 수의 측면에서 보면 전체 사업체의 약 0.1%인 1,289개소를 차지하고 있는 500인 이상 규모에서 전체 근로자의 12.9%를 차지하고 있다. 환경보호지출 관련 조사항목도 비슷한 경향을 보일 것으로 예상된다. 「환경보호지출계정」 조사는 그 속성상 환경보호경상지출액이나 인건비의 사업체 규모별 편차가 대단히 커서 주요 조사 변수에 대한 왜도가 대단히 큰 값을 갖게 된다. 이와 같은 경우에 현행 표본설계와 마찬가지로 전수조사층과 표본조사층을 구성하여 표본설계하는 것은 추정의 정확도를 높일 수 있는 효과적인 방안이다.

(4) 층화 경계점 결정

표본추출틀로 「경제총조사」 결과를 이용할 수 있다면 매출액 정보를 층화변수로 하는 Lavallée & Hidiroglou 방법을 적용하여 전수층과 표본층의 층 경계점을 구한다. 공표일정의 지연으로 「경제총조사」 결과를 표본추출틀로 활용할 수 없는 경우에는 「전국사업체조사」 결과를 이용하여 표본설계를 진행해야 한다.

「전국사업체조사」 결과에는 매출액 정보를 포함하고 있지 않기 때문에 사업체 규모를 나타내는 보조변수인 근로자 수 정보를 층화변수로 이용하여 Lavallée & Hidiroglou 방법을 적용한다.

Lavallée & Hidiroglou 방법(절사층화추출법)은 모수 추정에 대한 목표상대표

준오차와 표본 층에 대한 표본배분법(예: 떡배분이나 네이만배분 등)이 정해진 상황에서 전수층과 표본층을 구분하고, 표본층에 대해서 층화임의추출법을 적용할 때 목표상대표준오차를 만족하는 조건 하에서 표본크기를 최소화하는 층 경계점을 결정하기 위한 방법이다. 이 방법에 의한 전수층과 표본층의 층 경계점을 구하기 위한 반복계산법(iterative method)에 의한 알고리즘은 Lavallée & Hidioglou (1988)에 제시되어 있다.

Lavallée & Hidioglou 방법(절사층화추출법)을 적용할 때 층화변수는 앞서 제시한 바와 같이 매출액 정보나 근로자 수 정보를 이용하고, 각 업종 구분에 대해서 절사층화추출법을 적용한다. 각 업종 구분에 대한 세부 층의 수(L)는 각 업종 내 사업체 수를 고려하여 차등을 두되 전수층을 포함하여 4~6개로 정하는 것이 바람직할 것이다.

(5) 각 업종별 목표오차 설정

현행 표본설계에서는 산업은행의 공해방지투자규모 및 환경부의 「제조업체 환경오염방지활동 및 투자현황」을 근거로 8개 업종(종이제조업, 석유정제품, 화합물·화학제품, 고무·플라스틱, 비금속광물, 제1차 금속, 조립금속 및 기계장비제조업)을 환경오염산업으로 규정하여 상대적으로 더 많은 표본을 배정하여 조사하고 있다. 새로운 표본설계에서는 환경보호지출액과 관련하여 업종별 중요도가 다르다는 점을 고려하여 업종별 목표오차에 차등을 두어 표본크기를 결정하는 방안을 적용해야 할 것이다. 예를 들어 현행 표본설계와 마찬가지로 8개 업종(종이제조업, 석유정제품, 화합물·화학제품, 고무·플라스틱, 비금속광물, 제1차 금속, 조립금속 및 기계장비제조업)에 대해서는 목표오차를 낮추어 설정함으로써 상대적으로 표본크기를 늘리는 것이 모집단 총계를 추정할 때 효율적일 것이다.

(6) 표본크기 결정

「환경보호지출계정」 결과의 통계이용자들은 산업 세부 업종별 통계를 요구하고 있다. 한편, 앞선 <표 3>에 제시한 2009년 「환경보호지출계정」 조사데이터 분석결과를 보면 환경보호경상지출액이나 인건비의 업종별 편차가 대단히 큰 것을 알 수 있다. 따라서 층화에 사용된 업종 구분은 통계작성단위의 역할을 하게 된다. 따라서 각 업종의 중요도를 고려해 업종별 추정의 목표상대표준오차를 결정한 후 이를 기초로 업종별 표본크기를 결정하는 것이 바람직할 것이다. 전체 표본크기는 각 업종별 표본크기의 합계로 구한다.

각 업종 구분에 대해서 절사층화추출법을 적용할 경우에 각 업종 구분에서 표본크기는 다음 식을 통해서 구할 수 있다. 이 식에서 W_h , S_h^2 , \bar{Y}_h 등은 모두 층 경

계점 $b_{(1)}, b_{(2)}, \dots, b_{(L-1)}$ 의 함수이다.

$$n = N \cdot W_L + \frac{N \left(\sum_{h=1}^{L-1} W_h^2 S_h^2 / a_h \right)}{N (cv_0 \bar{Y})^2 + \sum_{h=1}^{L-1} W_h S_h^2}$$

여기서, L 은 전체 층의 수, $a_h = \frac{N_h^p S_h^q}{\sum_{h=1}^{L-1} N_h^p S_h^q}$, ($0 < p < \infty, 0 \leq q < \infty$), cv_0 는 목표상

대표준오차이다.

참고로 2009년 「환경보호지출계정」 조사데이터를 기초로 제조업의 업종별 목표 변동계수는 15%로 하고, 비제조업의 목표변동계수는 30%로 설정하여 간편 방법에 따라 구한 전체 표본크기는 4,120개소이다.

새로운 표본설계에서는 앞서 제시한 Lavallée & Hidiroglou 방법(절사층화추출법)에 의한 표본크기 공식을 이용하여 각 업종별 표본크기를 결정할 수 있으며, 사전정보가 부족한 상황이라면 <표 11>에 제시한 바와 같이 업종별 표본사업체 수를 결정한 후 Lavallée & Hidiroglou 방법을 이용하여 전수층과 표본층의 층 경계점을 구할 수 있을 것이다.

(7) 표본배분 및 표본추출

현행 표본설계에서 표본층에 대한 표본배정은 환경투자 및 비용과 상관관계가 높은 것으로 조사된 매출액을 대표 변량으로 하여 각 업종 내에서 매출액 규모로 세부 층화한 후에 네이만배분법에 따라 표본을 배분하였다.

새로운 표본설계에서 층화변수는 매출액이나 근로자 수 정보이다. 표본배분 단계에서는 매출액 정보를 이용하여 네이만배분법을 적용하는 방안을 고려할 수 있다. 매출액 정보를 이용할 수 없다면 근로자 수 정보를 이용하거나 각 세부 층별 사업체 수 기준의 역배분법(power allocation) 또는 제곱근비례배분법 등을 적용하는 방안도 고려할 수 있을 것이다. 최종 표본배분법은 이들 배분방법을 비교하여 결정하도록 한다.

각 층별로 표본크기가 결정되면 각 층에서는 사업체를 행정구역 구분(또는 근로자 수)에 따라 정렬한 후 계통추출법을 적용하여 표본사업체를 선정한다.

- 층의 경계점 결정과 표본배분 계산 예시 : 앞서 제시한 층화, 층 경계점 결정, 표본배분 등의 과정을 ‘음식료 제조업’과 ‘섬유, 의복, 모피 제조업’ 대상으로 살펴본 결과를 제시한다. 여기서는 앞서 제시한 두 업종 구분에 대해서 상시 근로자 1인 이상을 대상으로 하는 경우와 상시근로자 수가 5인 이상인 사업체를 대상으로 하는 경우에 대해 각각 층 경계점과 표본배분 결과를 제시한다.

- 업종별 표본크기는 앞선 <표 11>에서 결정된 업종별 표본크기를 이용하였음
 - 근로자 수를 층화변수로 가정하여 결정한 층 경계점 및 표본배분 결과임
 - 층 경계점은 Lavallée & Hidiroglou 방법을 적용하여 구한 결과임
 - 표본층 표본배분법은 네이만 배분법을 적용하여 구하였음
 - 각 업종별 층의 수는 5개인 경우와 6개인 경우를 함께 검토하였음
- ‘음식료 제조업’과 ‘섬유, 의복, 모피 제조업’의 층 경계점과 층별 표본배분 결과는 <표 12>와 <표 13>에 각각 제시하였다. 층수가 동일한 경우에 업종 구분에 따른 층 경계점은 비슷한 수치를 갖는 것을 살펴볼 수 있다. 업종별 층 경계점을 결정할 때는 각 업종별 층 경계점을 결정하는 방안과 전체 업종에 대해서 공통적으로 적용되는 층 경계점을 정하는 방안의 장단점을 비교하여 결정해야 할 것이다.

또한 분석 결과를 살펴보면 층의 수가 늘어날 때 추정의 정확도가 높아지고 있다. 따라서 새로운 표본설계에서 각 업종별 층수는 6개로 결정하는 것이 바람직할 것으로 판단된다. 다만, 모집단 크기와 표본크기가 지나치게 작은 경우는 층의 수를 줄여야 할 것이다.

층의 경계점의 구분이 기존의 규모층 구분과 상이하므로 유사통계와 비교분석하는데 일치성의 문제가 발생할 수 있으므로 이론적인 계산결과와 이용 상 편의성을 고려한 실제적인 내용을 함께 고려하여 층의 경계점과 배분방법을 선택하여 적용해야 생산되는 통계의 정확성과 활용성을 제고할 수 있을 것이다.

아울러 향후 매출액과 근로자 수에 대한 통계적 모형 적합이 가능하다면 Rivest (2002) 방법을 적용하여 층 경계점을 구하는 방안도 검토해야 할 것이다.

<표 12> 음식료 제조업(담배제조업 포함)에 대한 층 경계점 결정

층 수	층번호	근로자 1인 이상 사업체 대상			근로자 5인 이상 사업체 대상		
		층 구간	N_h	n_h	층 구간	N_h	n_h
5	1	7.5 이하	9,870	17	15.5 이하	4,322	17
	2	7.5-28.5	3,388	18	15.5-43.5	1,769	18
	3	28.5-96.5	1,214	20	43.5-117.5	689	18
	4	96.5-432.5	320	20	117.5-411.1	238	20
	5	432.5 이상	19	19	411.1 이상	21	21
	계		14,811	94		7,039	94
6	1	5.5 이하	8,574	13	12.5 이하	3,773	13
	2	5.5-18.5	3,909	16	12.5-28.5	1,713	13
	3	18.5-51.5	1,568	17	28.5-64.5	968	16
	4	51.5-132.5	545	14	64.5-148.5	413	16
	5	132.5-432.5	196	15	148.5-411.0	151	15
	6	432.5 이상	19	19	411.0 이상	21	21
	계		14,811	94		7039	94

Note: 업종별 근로자 수 추정에 대한 상대표준오차는 근로자 1인 이상 사업체를 대상으로 할 때 층의 수가 5와 6인 경우에 각각 4.5%와 3.6%를 나타냄. 근로자 5인 이상 사업체를 대상으로 할 때 층의 수가 5와 6인 경우 각각 3.4%와 2.7%

를 나타냄.

<표 13> 섬유, 의복, 모피 제조업에 대한 층 경계점 결정

층 수	층번호	근로자 1인 이상 사업체 대상			근로자 5인 이상 사업체 대상		
		층 구간	N_h	n_h	층 구간	N_h	n_h
5	1	6.5 이하	18,853	23	12.5 이하	7,526	19
	2	6.5-21.5	6,496	20	12.5-31.5	3,020	18
	3	21.5-76.5	1,966	21	31.5-86.5	1,104	18
	4	76.5-463.0	355	20	86.5-378.0	288	20
	5	463.0 이상	13	13	378.0 이상	22	22
	계		27,683	97		11,960	97
6	1	4.5 이하	15,723	16	10.5 이하	6,637	17
	2	4.5-12.5	7,526	16	10.5-23.5	3,220	17
	3	12.5-32.5	3,076	16	23.5-52.5	1,439	18
	4	32.5-94.5	1,091	17	52.5-127.5	501	15
	5	94.5-422.5	253	18	127.5-422.5	149	16
	6	422.5 이상	14	14	422.5 이상	14	14
	계		27,683	97		11,960	97

Note: 업종별 근로자 수 추정에 대한 상대표준오차는 근로자 1인 이상 사업체를 대상으로 할 때 층의 수가 5와 6인 경우에 각각 4.5%와 3.6%를 나타냄. 근로자 5인 이상 사업체를 대상으로 할 때 층의 수가 5와 6인 경우 각각 3.3%와 2.7%를 나타냄.

(8) 가중치 작성 및 추정방법

일반적으로 복합표본조사(complex sample survey)의 가중치는 ㉠ 설계 가중치, ㉡ 무응답에 대한 조정, ㉢ 사후층화에 대한 조정 등의 세 가지 요인을 통합하여 산정된다. 본 조사에서 표본사업체 가중값을 구하기 위해서는 설계 가중값과 무응답 조정의 과정을 거쳐야 한다. 우선 설계 가중값과 무응답 조정값을 구하여 표본 사업체에 대한 가중값을 구하고, 다음 단계로 벤치마킹 과정(사후층화 조정)을 거쳐야 할 것이다. 표본설계에서 사후층화 단계에서 사용되는 벤치마킹 정보는 활용할 수 있는 최신 정보의 「전국사업체조사」 결과의 근로자 수 정보나 「경제총조사」 결과의 매출액 정보를 이용한다. 모집단의 보조정보를 가중치 작성에 반영하여 모수를 추정할 수 있으며 정확한 보조정보가 있는 경우에 보조정보를 추정단계에서 이용하면 추정의 정도는 상당히 향상될 수 있다. 벤치마킹 과정(사후층화 조정)에서 사후층의 결정은 업종*규모 구분을 이용하는 방안과 업종 구분만을 이용하는 방안을 비교하여 최종 결정해야 할 것이다.

이 조사에서 주요 조사항목에 대한 모총계 및 모평균 추정은 가중치를 이용한 추정량을 사용하며, 모평균에 대한 추정량은 표본가중평균으로 아래 식으로 계산한다.

$$\hat{\mu}_h = \frac{\sum_{i=1}^{n_{hi}} \sum_{j=1}^{n_{ij}} w_{hij} y_{hij}}{\sum_{i=1}^{n_{hi}} \sum_{j=1}^{n_{ij}} w_{hij}}$$

여기서, h 는 업종을 나타내고 i 는 규모층을 나타내며 j 는 표본사업체를 의미하고 w_{hij} 와 y_{hij} 는 각각 h 업종 i 규모층 내의 j 사업체의 가중치와 관찰값을 나타낸다. 모총계에 대한 추정량은 다음 식으로 계산한다.

$$\hat{\tau}_h = \sum_{i=1}^{n_{hi}} \sum_{j=1}^{n_{ij}} w_{hij} y_{hij}$$

여기서, h 는 업종을 나타내고 i 는 규모층을 나타내며 j 는 표본사업체를 의미하고 w_{hij} 와 y_{hij} 는 각각 h 업종 i 규모층 내의 j 사업체의 가중치와 관찰값을 나타낸다.

표본조사 데이터를 분석할 때 가중치를 무시하고 분석하면 모수 추정에 심각한 편향(bias)이 발생할 수 있고, 추정량의 분산이 과소평가되어 문제가 될 수 있다. 새로운 표본설계에서는 각종 총계 추정량에 대한 표본설계를 반영한 분산추정량과 상대표준오차 산출식을 제공해야 할 것이며, 아울러 추정량의 상대표준오차를 복합표본조사 분석용 통계패키지인 SAS, SPSS 등을 이용할 수 있는 방안을 구체적으로 제시해야 할 것이다.

(9) 특이치 검출

「환경보호지출계정」 조사의 조사항목은 그 속성상 특이치가 빈번하게 발생할 가능성이 있다. 따라서 특이치를 검출하여 조사값을 점검하는 것은 조사의 정확도를 높이는 중요한 과정이다.

「환경보호지출계정」 조사에서는 해당년도와 전년도의 조사항목 값을 함께 조사하고 있다. 따라서 중요 조사항목에 대한 특이치 검출은 Hidiroglou and Berthelot 방법을 사용하는 것을 검토해야 할 것이다. 참고로 이의규(2010)는 광업 및 제조업 조사에 Hidiroglou and Berthelot 방법을 사용하여 특이치를 검출하였다. 이의규(2010)와 Hidiroglou and Berthelot(1986)을 참고하면 「환경보호지출계정」 조사를 위한 효율적인 특이치 검출방안을 마련할 수 있을 것이다.

(10) 모집단 및 표본 관리

표본조사를 통해 얻어지는 환경보호지출계정 통계의 품질을 높이기 위해서는 연구 대상 모집단의 특성과 시간의 흐름에 따른 변동을 반영할 수 있는 대표성 있는 표본 사업체를 대상으로 조사가 이루어져야 한다. 이를 위해서는 표본설계 뿐

만 아니라 지속적인 표본 사업체 관리가 필수적이다.

새로운 표본설계에서는 모집단 및 표본 관리 방안을 제시해야 한다. 예를 들어 표본으로 선정된 사업체도 폐업이나 전업 등으로 인해 변동이 생기고 또한 새로운 사업체가 창업할 경우 이들도 표본에 반영해 주어야 한다. 또한 표본추출틀 작성 시점과 조사 시점간의 차이를 줄이기 위해서는 매년 새로이 창업하는 사업체를 모집단에 반영하여 표본 사업체를 추가하고, 폐업, 휴업 또는 전업한 표본 사업체를 다른 사업체로 교체하는 등의 방법 등을 제시해야 할 것이다.

제 7 장 결언

환경보호지출계정은 환경보호지출에 대한 지표를 제공하고 국내환경산업의 생산 구조와 고용구조를 파악함과 동시에 국가 환경정책의 유효성을 평가하는데 사용되는 중요한 국가승인통계이다. 본 통계의 품질관리상태를 진단하기 위한 2010년 국가통계품질진단 결과에서 지적된 문제점인 '제조업과 비제조업간의 조사단위 불일치 문제 해소 필요'와 '환경보호지출계정 표본의 최신화 필요' 등에 대한 실태분석을 통해서 개선안을 마련하기 위하여 다음과 같은 내용을 연구하였다.

먼저 환경보호지출계정의 작성실태를 분석하였으며 여기서는 표본설계에 관한 세부적인 설명문서가 없어서 구체적으로 파악할 수 없었지만 2010년 품질진단에서 지적된 문제점 외에도 '가중치 작성과 모수추정'과 '특이치(outlier) 검색' 등에서 문제점을 파악하였다.

두 번째는 환경보호지출계정의 품질개선을 위해서 해결해야 할 사항을 분석하였다. 결과에서 도출된 세 가지 사안은 '표본추출틀의 변경', '표본설계 개선'과 '신규개선 표본설계와 기존 표본설계 간의 시계열유지 방안' 등이었다. 특히 시계열유지에서는 산업분류체계가 KSIC 8에서 KSIC 9로 변경되면서 업종간의 연계방안의 문제가 표출되었다.

세 번째는 업종분류체계 변경 방안을 검토하기 위해서 기존 데이터에 부여된 코드를 분석하여 새로운 분류체계와 연결하는 방안을 검토하였으나 기존 데이터에서 사업체의 분류코드를 알 수 없으므로 업종분류체계 변경에 따른 정확한 해결 방안을 찾기 어려웠다.

네 번째는 표본틀 변경 방안 연구로서 '제조업과 비제조업의 조사단위 단일화', '표본의 최신화'와 '조사대상에 대한 명확한 정의' 등이 연구되었다. 최신의 표본틀로 '2010년 경제총조사'가 제안되었으나 현시점에서 마이크로데이터 이용이 불가능하여 '2010년 전국사업체조사'를 이용하여 분석하였다.

마지막으로 최적표본설계방안을 표본추출 가이드라인의 형식으로 제시하였다. 일반적인 표본설계 기본원칙을 설명한 후에 신규표본설계에서 정확한 추정을 위해 필요한 변형절사추출법의 적용방법을 제시하였다. 층화변수로서 생산액이나 매출액을 제안하였으나 전국사업체조사에서는 제공되지 않으므로 종사자수를 대신 사용하였고 표본크기 결정은 2009년 EPEA 데이터를 기준으로 목표변동계수를 설정하여 산출하는 방법을 제시하여 전체 표본크기를 4,120개소로 제안하였다.

다음에는 다양한 특이치 검색방법을 제시하여 조사항목의 특성별로 적합한 특이치 검색방법을 적용할 수 있도록 하였고 표본설계에 따라 적절한 가중치 산출방법과 모수추정법을 설명하였다.

표본설계의 가이드라인 설명을 위해서 ‘2013년 조사를 위한 표본설계 연구절차’를 10개 과정으로 구분하여 제시하였다.

‘환경보호지출계정’의 품질개선을 위해서 취할 수 있는 ACTION PLAN인 가이드라인을 제시하였으나 ‘2010년 경제총조사’ 데이터를 직접 활용하지 못하여 실질적이고 수치적인 예시를 보이지 못한 한계는 있지만 6.7절에 주어진 ‘2013년 조사를 위한 표본설계 연구절차’를 적용한다면 2013년 이후에 생산되는 ‘환경보호지출계정’의 통계품질은 획기적으로 개선될 것으로 기대한다.

2009년 ‘환경보호지출계정’의 조사데이터로부터 통계생산까지 과정에 대한 내용이 문서화되지 않아서 전체적인 내용을 이해하는데 한계가 있었으므로 언젠가는 반드시 표본설계 내용을 포함한 ‘환경보호지출계정’의 전체 과정을 문서화했으면 하는 바램이며 본 연구를 위한 자료제공 등의 적극적인 협조와 지원을 해주신 통계청 품질관리 담당자분들과 한국환경정책·평가연구원의 실무자분들에게 감사 말씀을 드리며 본연구가 ‘환경보호지출계정’의 품질개선에 도움이 되기를 바랍니다.

부록 1. 다양한 변동계수 크기에 따른 표본크기 산출결과

<부록 1> 인건비와 환경보호경상지출 대한 필요 표본크기

업종 중분류	표본 크기	인건비					환경보호경상지출					
		10%	15%	20%	25%	30%	10%	15%	20%	25%	30%	
음식류	141	98	44	25	16	11	167	74	42	27	19	
섬유의복모피	104	219	97	55	35	24	456	203	114	73	51	
목재나무	43	156	69	39	25	17	528	235	132	84	59	
종이인쇄 출판	필프종이	70	100	44	25	16	11	148	66	37	24	16
	출판인쇄	97	697	310	174	112	77	367	163	92	59	41
석유정제 화학고무	코크스 석유정제	26	359	159	90	57	40	569	253	142	91	63
	화합물 화학	177	417	185	104	67	46	446	198	112	71	50
	고무 플라스틱	151	1315	584	329	210	146	949	422	237	152	105
비금속광물	148	245	109	61	39	27	726	323	182	116	81	
제1차금속	130	631	280	158	101	70	4327	1923	1082	692	481	
조립금속 기계장비	조립금속	139	225	100	56	36	25	314	140	79	50	35
	기계장비	211	633	281	158	101	70	2406	1069	601	385	267
운수장비	99	392	174	98	63	44	481	214	120	77	53	
가구기타	122	355	158	89	57	39	521	232	130	83	58	
축산재배임업	61	71	32	18	11	8	368	163	92	59	41	
어업	16	25	11	6	4	3	74	33	19	12	8	
광업	32	1419	631	355	227	158	827	367	207	132	92	
전기가스수도	40	346	154	86	55	38	353	157	88	56	39	
건설업	103	1046	465	262	167	116	1999	888	500	320	222	
도소매업	161	2963	1317	741	474	329	2794	1242	699	447	310	
음식숙박	78	1069	475	267	171	119	1507	670	377	241	167	
운수창고 통신	운수창고	64	946	420	236	151	105	1458	648	364	233	162
	통신	11	0	0	0	0	0	146	65	36	23	16
금융부동산임대	98	2914	1295	729	466	324	1764	784	441	282	196	
보건	102	161	72	40	26	18	396	176	99	63	44	
기타서비스업	95	479	213	120	77	53	1925	856	481	308	214	
합 계	2,519	17,281	7,679	4,321	2,764	1,918	26,016	11,564	6,505	4,160	2,890	

- 표본크기 산출공식 : $n_{2012} = n_{2009} \cdot \left(\frac{CV_{2009}}{CV_{2012}} \right)^2$

<참고문헌>

- [1] 박인호 (2007), 절사표본조사에 관한 연구, 국민계정(3), 81-106.
- [2] 이광한 (2008) 환경보호지출계정(EPEA) 개발결과, 국민계정(4), 88-136
- [3] 이의규 (2010) 주기적 조사 자료의 내검 : 그래프 활용을 중심으로, 통계연구 15, 16-27.
- [4] Bee, M., Benedetti, R., and Espa, G. (2007), "A Framework for Cut-off Sampling in Business Survey Design", University of DeGli Studi - Department of Economic, working paper.
- [5] Clark, K.K. and Kinyon, D.L. (2007), "Can We Continue to Exclude Small Single-Establishment Businesses from Data Collection in the Annual Retail Trade Survey and the Service Annual Survey?" presented at the Third International Conference on the Establishment Surveys in Montreal, Quebec, Canada.
- [6] Elisson, H. and Elvers, E. (2001), "Cut-off Sampling and Estimation," Proceedings of Statistics Canada Symposium on Achieving Data Quality in a Statistical Agency: A Methodological Perspective.
- [7] Hidioglou, M. A. and Berthelot, J. M. (1986) Statistical edit and imputation for periodic surveys, Survey Methodology 12, 73-83.
- [8] Lavallée, P., and Hidioglou, M. A. (1988) On the stratification of skewed populations. Survey Methodology 14, 33-43.
- [9] Rivest, L-P. (2002), A Generalization of the Lavallée and Hidioglou Algorithm for Stratification in Business Surveys. Survey Methodology, 28, 191-198.
- [10] Rouseeuw and Zomeren (1990) Unmasking multivariate outliers and leverage points, Journal of the American Statistical Association 85, 633-639.
- [11] Sethi, V. K. (1963). A note on the optimum stratification of populations for estimating the population means. Australian Journal of Statistics, 5, 20-23.
- [12] Thompson, K. J. (2007) Investigation of Macro editing techniques for outlier detection in survey data, proceedings of the third international conference on establishment surveys, American Statistical Association.