

## 연구논문

## 노동부 통계자료에 대한 새로운 조사방법의 영향 평가

Influence Appraisal for Labor Statistics after Introducing a New Survey Method

성 내 경\*  
Nae Kyung Sung

노동부에서 시행하는 《노동력수요동향조사》와 《임금구조기본통계조사》를 기초로 새로운 조사표와 조사방법의 도입으로 인한 영향을 분석하고 평가하는 간편한 통계적 방법을 제안한다. 두 조사 모두 매우 유사한 표집틀(sampling frame)과 표집법을 공유하고 있기 때문에 두 조사 중 어느 하나를 기준으로 다른 하나의 변동성을 측정할 수 있다. 여기서 제시하는 영향 평가는 연도별로 동일 범주에 속하는 두 독립적인 추계치들의 비를 기초로 진행되며 또한 새로운 조사의 도입 시점을 기준으로 전과 후로 분할된 비 데이터(ratio data)의 변동성을 통계적으로 비교하는 형식을 갖는다.

주제어: 영향 평가, 노동통계, 비 데이터, 외부오차

Based upon the Labor Demand Survey and the Survey on Wage Structure being conducted annually by the Ministry of Labor, we suggest a convenient statistical tool which can analyze and appraise the effect of introducing a new survey method. Since both surveys have in common very similar sampling frames and sampling schemes in structure, one can measure the variability of one survey on the basis of the other. The influence appraisal method adopted here is applied to ratio data between two independent estimates belonging to the identical category by year and has a statistical form of comparing ratio data before introducing a new survey with those after introducing a new survey.

key words: influence appraisal, labor statistics, ratio data, external error

\* 교신저자(corresponding author): 고려대학교 통계연구소 특별연구원 성내경.  
E-mail: hyeilsnk@hanafos.com

## I. 개요

노동부는 노동력 수급상황 파악과 고용증감 전망 등 인력정책 수립에  
서 기초자료로 활용하기 위해 매년 4월~5월에 《노동력수요동향조사》  
를 실시한다. 조사대상은 전국의 상용근로자 5인 이상을 고용하고 있는  
사업체에서 표본추출한 13,800개 사업체들이다(노동부 2001). 표본은 고  
용규모별로 층화계통법에 의하여 추출한다. 조사방식은 타계식 면접조사  
며, 조사내용은 사업체별 근로자의 직종별, 산업별, 사업체 규모별 현재  
인원과 부족 인원이다.

이 조사는 원래 《고용전망조사》라는 이름으로 1976년에 처음 시행되  
었고 1995년부터 현재의 명칭으로 바뀌었지만 1998년까지는 조사내용과  
조사방법에 큰 변화는 없었다. 그러나 1990년대 중반 이후 노동시장의 팽  
창과 다양한 직업분화 현상이 두드러지면서 노동력 수요에 대해 더욱 상  
세한 파악이 필요해졌다(어수봉 1994, 이병직 1995, 한진수·최강식  
1997, 홍두승 1997, 남성일 1998, 김수곤·윤석천 1998). 이에 노동부는  
《노동력 수요동향조사의 표본설계 및 조사표 개발에 대한 연구》(동국대  
학교 자연과학연구원 1998)를 실시하였고 이 결과 새로운 조사표와 조사  
방법을 도입하여 현재에 이르고 있다. 지금까지 1999년, 2000년, 2001년  
3차에 걸쳐 새로운 조사표와 조사방법에 의거하여 조사가 진행되었다. 새  
로운 조사방법의 주요 내역은 자계식을 타계식으로 전환하고, 조사대상을  
상용근로자 10인 이상 사업체에서 상용근로자 5인 이상 사업체로 확대하  
고, 산업대분류·직업소분류 수준에서 집계하던 근로자수를 산업소분류·  
직업세분류 수준에서 집계하는 것이다.

그러나 이러한 조사대상과 조사방법의 변화에 따른 노동력 수요조사 결  
과에 대한 평가는 없었다(김병조·이건 2001). 그렇다면 1998년까지의 조  
사와 1999년부터 그 이후의 조사결과를 상호비교하여 조사방법의 변경에  
따른 차이 또는 새로운 조사방법의 도입으로 인한 상대적 변화를 정량적으

로 평가할 수 있는 방법이 있을까? 만일 이것이 통계적으로 가능하다면 향후 《노동력수요동향조사》의 보완과 개선에 일조할 수 있을 것이다.

그러나 물론 《노동력수요동향조사》 자료만으로 이러한 영향 평가는 불가능하다. 동등히 비교 가능한 기준 자료가 존재해야한다. 다행히 노동부 노동경제담당관실에서는 《노동력수요동향조사》와 더불어 《임금구조기본통계조사》를 실시하고 있는 바 두 조사의 표집틀은 거의 동일하며 또한 《노동력수요동향조사》의 주요 변수인 산업별·직종별 근로자수를 《임금구조기본통계조사》에서도 추계하고 있다.

《임금구조기본통계조사》는 근로자의 임금과 근로시간 등 제반 근로조건에 관련된 사항들을 산업별, 직종별로 파악하여 정부의 경제정책과 임금정책을 위한 기초자료 마련을 위하여 매년 6월을 기준으로 시행하고 그 결과를 다음 해 8월에 발표하는 자계식 표본조사로서, 전국의 5인 이상 사업체 중 고용규모별로 층화계통법에 의하여 추출된 5,400개 표본사업체를 대상으로 한다. 참고로 《임금구조기본통계조사》에서도 1999년 이전에는 《노동력수요동향조사》의 경우와 마찬가지로 전국의 10인 이상 사업체가 표집대상이었다.

두 조사 모두 동일 부서에서 관장하며 매년 비슷한 시기에 거의 동일한 표집틀을 기초로 매우 유사한 표본추출법을 사용하기 때문에 두 조사 결과의 일치성은 상당히 높을 것으로 기대된다. 그리고 《노동력수요동향조사》의 조사표와 조사방법이 변경된 1999년을 전후하여 《임금구조기본통계조사》의 조사표와 조사방법은 표집대상의 확대를 제외하면 크게 변경된 사실이 없기 때문에 《임금구조기본통계조사》는 내부적 일관성을 유지한다고 판단할 수 있다. 따라서 《임금구조기본통계조사》에서 추계한 산업별·직종별 근로자수는 새로운 조사방법이 도입된 《노동력수요동향조사》의 상대적 영향과 변동을 측정하는 유용한 기준이 될 수 있다. 즉 두 조사에서 공통적으로 조사하는 산업·직업별 근로자수의 연도별 추이를 분석하면 새로운 조사방법으로 인한 차이점을 파악하는 한 가지 단서를 얻을 수 있을 것이다.

연도별로 발간되는 《노동력수요동향조사 보고서》와 《임금구조기본 통계조사 보고서》에 수록된 산업·직업별 근로자수에는 다소간 차이가 존재한다. 물론 표집오차(sampling error)로 인한 차이는 자연스러운 것이지만 새로운 조사방법의 도입 시기를 전후하여 연도별로 산업·직업별 근로자수의 차이에 통계적으로 유의한 패턴이 존재한다면 이러한 부분을 새로운 조사표와 조사방법의 도입에 따른 영향으로 해석해도 무방할 것이다. 그러나 반대로 새로운 조사방법의 도입으로 인한 변동이 존재하지 않는다면 두 조사에서 추계한 산업·직업별 근로자수의 차이는 단순한 오차 변동에 불과하다.

## II. 가용 데이터와 모형

2003년 9월 현재 노동부 웹사이트(<http://www.molab.go.kr>)에서 제공하는 가장 최근에 발간된 《2002년 노동력수요동향조사 보고서》와 《2001년 임금구조기본통계조사 보고서》를 기초로 하여 차후 분석에 공통적으로 가용한 연도, 산업대분류, 직업대분류를 <표 1>에 요약하였다.

<표 1> 두 보고서에서 공통적으로 가용한 연도, 산업대분류, 직업대분류

연도	산업대분류	직종대분류
1994	C. 광업	고위임직원 및 관리자
1995	D. 제조업	전문가
1996	E. 전기·가스 및 수도사업	기술공 및 준전문가
1997	F. 건설업	사무직원
1998	G. 도소매 및 소비자용품 수리업	서비스 및 판매근로자
1999	H. 숙박 및 음식점업	농업 및 어업 숙련근로자
2000	I. 운수·창고 및 통신업	기능원 및 관련 기능근로자
2001	J. 금융 및 보험업	장치, 기계조작 및 조립원
	K. 부동산 임대 및 사업서비스업	단순노무직 근로자
	M. 교육서비스업	
	N. 보건 및 사회복지사업	
	O. 기타 공공 사회 및 개인서비스업	

단, 《임금구조기본통계조사》 데이터는 1993년부터 가용하며, 《임금구조기본통계조사》에서는 산업대분류로 농림업 및 수렵업, 어업의 두 범주를 추가하고 있으며, 또 직종대분류 중 서비스 및 판매근로자를 서비스근로자와 판매근로자의 두 항목으로 세분하고 있다. 또한 표에서는 《노동력수요동향조사》의 조사표와 조사방법이 바뀐 1999년부터 2001년까지를 강조하기 위하여 해당 연도를 검은 상자로 둘러 따로 표시하였다.

현재의 상황을 일반화하자.  $\tau_{ijk}$ 를  $i$  연도에  $j$  산업과  $k$  직업에 속한 근로자수라고 하자. 그리고 《노동력수요동향조사》에서 추계한  $\tau_{ijk}$ 의 추정값을  $l_{ijk}$ , 《임금구조기본통계조사》에서 추계한  $\tau_{ijk}$ 의 추정값을  $w_{ijk}$ 라 하자. 만일 두 조사의 추정값들이 통계적으로 일치한다면 대응하는 두 추정값들의 비(ratio)  $r_{ijk}$ 는 대략 1이 될 것이다. 덧붙여 동일 모수를 거의 동등한 두 가지 방법으로 추정하는 현재의 경우 두 추정량  $l_{ijk}$ 와  $w_{ijk}$ 의 값의 차이는 0을 중심으로 대칭적 분포를 이루며 또한 그 비의 분포는 1 근방에서 높은 확률밀도를 갖는다고 봄이 타당하다. 따라서 이 경우에 적합한 모형은 다음과 같다.

$$r_{ijk} = \frac{l_{ijk}}{w_{ijk}} = 1 + \varepsilon_{ijk}, \quad \varepsilon_{ijk} \sim S(0, \sigma_{jk}^2).$$

모형에서  $\varepsilon_{ijk}$ 는  $(-1, \infty)$ 에서 정의되는 오차항이다. 랜덤오차  $\varepsilon_{ijk}$ 는 평균 0을 중심으로 상당히 좁은 구간에서 정규분포와 매우 유사한 단봉의 거의 대칭인 연속분포  $S$ 를 따른다고 가정한다.  $\sigma_{jk}^2$ 은  $j$  산업과  $k$  직업에서 오차분산이다. 이 경우 우리의 주 관심변수인 두 추정값들의 비  $r_{ijk}$ 의 분포는  $S(1, \sigma^2)$ 가 된다. 그런데 두 추정량  $l_{ijk}$ 와  $w_{ijk}$ 의 분포는 기실 다항분포를 따르며 따라서  $r_{ijk}$ 의 분포는 수리적으로 두 다항확률변수들간 비의 분포로 정의된다. 또한  $r_{ijk}$ 는 항상 0보다 커야 하므로 통상적 오차분포인 정규분포 가정이 불가능하다. 다만 현재의 상황에서 두 추정값의 비는 언제나 1 근방에 위치할 것이 확실하므로 실용성의 측면에서  $r_{ijk}$ 의 분포로 정규분포의 성질을 그대로 유지하는 유사정규분포(pseudo-normal distribution)

를 가정하는 것이다. 즉 신뢰구간 계산이나 검증시 정규분포에 기초한 제반 방법론이 그대로 적용된다.

그런데 모형에서 비의 분자와 분모는 바뀌어도 무방하다. 다시 말해서 비를 다음과 같이 정의하고 분석을 진행해도 우리의 주관심사는 상대적 변동이기 때문에 결과는 같다.

$$r'_{ijk} = \frac{w_{ijk}}{l_{ijk}}$$

우리의 목표는 일정 시점  $t$ 를 전후하여  $w_{ijk}$ 를 기준으로  $l_{ijk}$ 의 값의 변화에 대한 통계적 유의성을 검증하는 것이다. 모형을 기준으로 말을 바꾸면, 일정 시점  $t$ 를 전후하여  $\varepsilon_{ijk}$ 가 현실화되는 패턴의 유의성을 검증하는 것이다.

$j$  산업과  $k$  직업에 속한 비 데이터(ratio data)를 다음과 같이 표기하자.

$$\{r_{1jk}, r_{2jk}, \dots, r_{tjk}, r_{(t+1)jk}, \dots\}$$

여기서 시점  $t$  이하에서는 과거식의 조사가 진행되었고  $(t+1)$  시점부터 새로운 조사방법이 도입되었다고 간주한다. 만일 새로운 조사방법의 도입 전후에 추정값 분포에 아무 차이도 없었다면  $r_{ijk}$ 의 분포 또는 동등하게  $\varepsilon_{ijk}$ 의 분포에는 전혀 변화가 없을 것이다. 반대로 조사방법의 도입으로 변화가 생겼다면  $(t+1)$  시점부터 나타나는  $\varepsilon_{ijk}$ 의 분포는 과거의 분포에 비하여 사뭇 다를 것이다.

따라서 새로운 조사방법의 도입 이전 데이터로부터 오차를 수량화하고 이 값을 기준으로  $(t+1)$  이상의 시점 데이터의 상대적 위치를 통계적으로 산정하면 새로운 조사방법의 도입으로 인한 상대적 오차변동의 효과를 탐색할 수 있다. 일종의 외부오차항(external error term)을 이용하여 분석하는 것이다. 랜덤오차의 추정량인 추정오차를 잔차(residual)  $e_{ijk}$ 로 표기한다. 즉,

$$e_{ijk} = \hat{\varepsilon}_{ijk}$$

오차분산  $\sigma_{jk}^2$ 의 추정량은 시점  $t$  이하의 비 데이터만으로 다음과 같이 계산한다.

$$s_{jk} = \hat{\sigma}_{jk}^2 = \frac{\sum_{i=1}^t (r_{ijk} - \bar{r}_{jk})^2}{(t-1)}$$

여기서 평균비(mean ratio)  $\bar{r}_{jk}$ 는 다음과 같이 정의한다.

$$\bar{r}_{jk} = \sum_{i=1}^t r_{ijk} / t$$

$(t+1)$  이상의 시점에서 계산된 잔차들에 대한 이상 여부를 판정하는 한 가지 쉬운 방법은 시점  $t$  이하의 잔차 데이터를 기초로 다음과 같이 비  $r_{ijk}$ 들의 모평균  $\rho_{jk}$ 에 대한 통상적인  $100(1-\alpha)\%$  신뢰구간을 구축하고 구간 외부에 존재하는 데이터를 유의하다고 판정하는 것이다.

$$\rho_{jk} \text{에 대한 } 100(1-\alpha)\% \text{ 신뢰구간: } \bar{r}_{jk} \pm \frac{s_{jk}}{\sqrt{t}} t_{t-1, \alpha/2}$$

이상과 같은 관점의 분석방식은 Belsley 등(1998)이 제안한 이래 회귀 모형에서 영향분석을 진행할 때 자주 사용된다. 기타 이러한 류의 분석시 비모수 통계기법인 부호검증, 절대잔차(absolute residual)를 이용한 검증법의 활용, 각종 잔차도 기법도 가능함을 첨언한다.

### III. 사례분석

제 II 장에서는 원래 주어진 문제와 상황에 수리적으로 접근하기 위하여 일반적으로 논리를 전개하였다. 그러나  $i$  연도에  $j$  산업과  $k$  직업에 속

한 근로자수  $\tau_{ijk}$ 에 대한 분석사례를 제시하기에는 예시할 데이터의 양이 만만치 않다. 따라서 여기서는 편의상 직업은 무시하고 오직 「산업별 근로자수 추계에서 새로운 조사방법의 도입에 따른 영향 평가」에 집중한다. 분석 개념은 동일하다.

〈표 2〉 《임금구조기본통계조사》에서 추계한 연도별 산업별 근로자수

연도 산업	1994	1995	1996	1997	1998	1999	2000	2001
전산업	4,923,705	5,219,747	5,301,550	5,249,178	4,729,350	5,469,934	5,735,072	6,150,915
C	28,125	26,577	22,803	20,260	16,655	18,864	19,140	16,346
D	2,530,918	2,638,565	2,550,291	2,397,382	2,021,143	2,257,528	2,389,136	2,580,969
E	35,447	41,773	44,243	46,122	47,876	44,833	46,104	45,412
F	279,570	283,088	291,503	301,147	264,736	281,253	298,349	302,147
G	365,380	415,077	441,652	455,583	369,950	602,368	653,355	712,983
H	76,272	82,592	81,346	81,703	74,780	118,611	148,875	164,883
I	505,638	510,720	530,775	537,078	515,558	488,186	489,245	505,258
J	340,194	352,128	364,729	358,273	332,668	440,235	440,137	401,543
K	382,891	432,081	499,021	552,967	569,884	536,923	541,848	652,943
M	138,170	159,477	169,974	177,877	191,946	257,023	285,486	295,943
N	150,378	164,725	188,203	198,936	201,943	230,039	234,643	272,048
O	90,722	112,944	117,010	121,850	122,211	158,353	156,942	177,094

〈표 2〉와 〈표 3〉은 각각 《임금구조기본통계조사》와 《노동력수요동향조사》에서 추계한 연도별 산업별 근로자수 데이터다. 전과 마찬가지로 《노동력수요동향조사》 데이터에서 새로운 조사방법이 도입된 1999년, 2000년, 2001년의 3개 년도에는 검은 상자선을 돌려 따로 표시하였다.

제 II 장에서 도입한 표기와 일관성을 유지하기 위하여 다음과 같은 표기를 약속한다.  $\tau_{ij}$ 를  $i$  연도에  $j$  산업에 속한 근로자수라고 하자. 즉,

$$\tau_{ij} = \sum_k \tau_{ijk}$$

〈표 3〉 《노동력수요동향조사》에서 추계한 연도별 산업별 근로자수

연도 산업	1994	1995	1996	1997	1998	1999	2000	2001
전산업	4,903,528	4,929,144	5,217,993	5,293,678	4,892,192	5,640,065	5,722,483	5,925,664
C	29,387	24,616	26,723	22,789	19,252	21,954	21,014	17,768
D	2,589,777	2,556,997	2,643,801	2,546,046	2,175,414	2,302,006	2,378,779	2,509,573
E	37,784	35,293	41,685	44,017	45,179	45,738	45,201	46,247
F	258,786	283,262	283,801	297,026	288,751	321,360	291,296	289,927
G	358,075	360,804	407,675	435,654	410,190	631,728	620,715	611,794
H	77,160	71,951	78,156	79,588	69,282	125,859	122,419	147,775
I	499,634	502,100	513,493	534,960	521,755	503,115	529,519	515,552
J	327,128	337,390	351,207	363,646	338,634	503,667	457,286	440,658
K	360,722	378,001	431,161	496,961	529,261	561,505	588,669	596,549
M	138,954	139,515	160,885	170,308	184,703	243,989	275,011	277,383
N	137,577	154,407	167,776	189,837	200,486	215,006	232,906	241,392
O	88,544	84,808	111,630	112,846	109,285	164,138	159,670	231,045

또 《노동력수요동향조사》에서 추계한  $\tau_{ij}$ 의 추정값을  $l_{ij}$ , 《임금구조 기본통계조사》에서 추계한  $\tau_{ij}$ 의 추정값을  $w_{ij}$ 라 하자. 이 사례분석에서 두 추정량들의 비  $r_{ij}$ 는 다음과 같이 정의하였다.

$$r_{ij} = \frac{w_{ij}}{l_{ij}}$$

이와 같이 계산한 비 데이터(ratio data)가 〈표 4〉에 주어져 있다.

〈표 4〉를 보면 1998년까지의 데이터의 경우 1995년에 두 조사의 차이가 가장 컸음을 알 수 있다. 특히 전산업 항목을 기준으로 볼 때 《임금구조기본통계조사》에서 추계한 근로자수가 《노동력수요동향조사》에서 추계한 숫자보다 5.9% 더 크다. 물론 이런 경우에 어느 조사에서 제시한 수치가 올바른지 알 방법은 없다. 다만 차이가 컸다는 사실만을 알 수 있다. 또 1998년도에도 두 보고서 간에 수치 차이가 많았음을 알 수 있다.

〈표 4〉 비 데이터 {  $r_{ij}$  }

연도 산업	1994	1995	1996	1997	1998	1999	2000	2001
전산업	1.004	1.059	1.016	0.992	0.967	0.970	1.002	1.038
C	0.957	1.080	0.853	0.889	0.865	0.859	0.911	0.920
D	0.977	1.032	0.965	0.942	0.929	0.980	1.004	1.028
E	0.938	1.184	1.061	1.048	1.060	0.980	1.020	0.982
F	1.080	0.999	1.027	1.014	0.917	0.875	1.024	1.042
G	1.020	1.150	1.083	1.046	0.902	0.954	1.053	1.165
H	0.988	1.148	1.041	1.027	1.079	0.942	1.216	1.116
I	1.012	1.017	1.034	1.004	0.988	0.970	0.924	0.980
J	1.040	1.044	1.039	0.985	0.982	0.874	0.963	0.911
K	1.061	1.143	1.157	1.113	1.077	0.956	0.920	1.095
M	0.994	1.143	1.056	1.044	1.039	1.053	1.038	1.067
N	1.093	1.067	1.122	1.048	1.007	1.070	1.007	1.127
O	1.025	1.332	1.048	1.080	1.118	0.965	0.983	0.766

〈표 5〉의 평균비(mean ratio)  $\bar{r}_j$ 는 새로운 조사방법이 도입되기 이전인 시점  $t$  이하의 비 데이터로부터 산출한 것이다.

$$\bar{r}_j = \sum_{i=1}^t r_{ij} / t$$

현재의 경우에 평균비  $\bar{r}_j$ 는 1994년부터 1998년까지 5개년 데이터를 이용하여 산출한다.

$$\bar{r}_j = \sum_{i=1994}^{1998} r_{ij} / 5$$

주어진 데이터에서 산업별 가용 연도의 개수는 5개에 불과하다. 관측수가 작기 때문에 신뢰구간 구축시 유의수준을 통상적인 유의수준 0.05보다 다소 높게 잡을 필요가 있다. 단, 두 조사방법 간에 큰 차이는 없기 때문에 지나치게 큰 유의수준은 불필요하다. 이같은 관점에서 현 데이터

분석시 유의수준은 0.1로 고정시켰다. 그리고 이 경우 소표본이기 때문에 90% 신뢰구간은  $t$  분포를 기초로 구축한다.

〈표 5〉 평균비와 90% 신뢰구간

1995년 데이터 포함		산업	1995년 데이터 제거	
평균비	90% 신뢰구간		평균비	90% 신뢰구간
1.007	(0.975, 1.040)	전산업	0.995	(0.970, 1.019)
0.929	(0.840, 1.018)	C	0.891	(0.837, 0.946)
0.969	(0.931, 1.007)	D	0.953	(0.927, 0.979)
1.058	(0.975, 1.141)	E	1.027	(0.959, 1.097)
1.008	(0.951, 1.064)	F	1.010	(0.929, 1.090)
1.040	(0.953, 1.128)	G	1.013	(0.921, 1.105)
1.057	(0.999, 1.114)	H	1.034	(0.990, 1.078)
1.011	(0.995, 1.027)	I	1.009	(0.987, 1.032)
1.018	(0.988, 1.048)	J	1.012	(0.974, 1.049)
1.110	(1.071, 1.150)	K	1.102	(1.052, 1.152)
1.056	(1.004, 1.107)	M	1.034	(1.002, 1.066)
1.067	(1.026, 1.109)	N	1.067	(1.008, 1.127)
1.121	(1.003, 1.238)	O	1.068	(1.020, 1.115)

〈표 5〉의 좌측 부분에 나와 있는 평균비와 90% 신뢰구간은 1994년부터 1998년까지 5개년 데이터를 이용하여 계산한 것이고, 우측에 나와 있는 평균비와 90% 신뢰구간은 1995년 데이터를 제외한 나머지 4개년 데이터를 기초로 산출한 것이다. 전술하였듯이 1995년 데이터는 두 조사 간의 산업별 추계 차이가 지나치게 크기 때문에 분산을 안정화시키기 위하여 제거하였다. 이와 같은 특이점 제거방법은 회귀분석시 영향진단기법으로 자주 사용되는 것이다.

현 사례분석에서는 특이한 데이터도 포함하는 5개년 데이터와 특이한 데이터 하나를 제거한 4개년 데이터를 기반으로 하는 두 가지 방법으로 데이터를 분석하는 바 연도수가 많으면 이런 경우에는 특이 데이터를 제거한 분석 하나만을 제시해도 무방하다. 그러나 현재 우리 데이터에는 가

용 관측이 5개밖에 되지 않는데 여기서 다시 자유도 하나를 더 상실한 채로 분석을 진행함은 바람직하지 않기 때문에 두 방법 모두를 동원해서 분석 결과의 공통성과 이질성을 관측해 볼 필요가 있다.

<표 6>에서는 5개년 데이터를 기준으로 연도별 추계 차이가 통계적으로 유의한 산업부문을 ●으로 표시하였다. 1999년부터 2001년까지를 보면 3년간 연속으로 추계 차이가 통계적으로 유의한 산업 범주들이 많음을 알 수 있다. 그러나 1995년과 1998년에도 유의한 칸이 많이 나타나고 있기 때문에 과연 최근 3년간의 유의성이 새로운 추세로 굳어진 것인지 아니면 전과 동일한 우연한 현상인지 직관적으로 판단하기 어렵다.

<표 6> 연도별 추계 차이가 유의한 산업 - 5개년 데이터 기준

산업 \ 연도	1994	1995	1996	1997	1998	1999	2000	2001
전산업		●			●	●		
C		●						
D		●			●			●
E	●	●						
F	●				●	●		
G		●			●			●
H	●	●				●	●	●
I			●		●	●	●	●
J				●	●	●	●	●
K	●		●			●	●	
M	●	●						
N			●		●		●	●
O		●				●	●	●

<표 7>은 1995년 데이터를 제거하고 나머지 4개년 데이터를 기초로 구축한 것으로 유의한 산업부문 칸을 검게 표시하였다. <표 7>을 보면 새로운 조사방법의 도입 이후 두 조사 간의 차이가 매우 커졌음을 확연히 알 수 있다. 1999년부터 2001년까지의 3개년 간 아무런 유의성도 검출되지

않은 산업범주 C와 E는 <표 2> 또는 <표 3>에서 확인할 수 있지만 그에 속하는 근로자수가 상대적으로 매우 적은 부문이다. 결국 근로자수가 아주 적지 않은 거의 모든 산업부문에서 《임금구조기본통계조사》와 《노동력수요동향조사》 간 추계 차이가 실재함이 극명하게 드러난다. 그리고 2000년도에 시행한 《임금구조기본통계조사》와 《노동력수요동향조사》에서 전산업 부문에 대한 추계는 둘 사이에 대차가 없었지만 내부적으로는 실제 산업분류 간에 추계 차이가 많았음을 알 수 있다.

<표 7> 연도별 추계 차이가 유의한 산업 - 4개년 데이터 기준(1995년 제거)

연도 산업	1994	1996	1997	1998	1999	2000	2001
전산업							
C							
D							
E							
F							
G							
H							
I							
J							
K							
M							
N							
O							

이상과 같이 외부오차를 기반으로 신뢰구간을 구축하여 유의성을 검증하는 방식 외에 새로운 조사표와 조사방법의 도입으로 야기된 영향을 평가하기 위하여 비평균과의 차이값 데이터를 활용하는 방안도 모색할 수 있다. 여기서는 개략만 제시한다.

<표 8>에는 모든 칸의 비 데이터에 대하여 1995년을 제외한 4개년 데

이터를 기준으로 산출한 비평균과 차이값 데이터  $\{r_{ij} - \bar{r}_j\}$ 가 주어져 있다. 여기서,

$$\bar{r}_j = \sum_i r_{ij} / 4, \quad (i = 1994, 1996, 1997, 1998)$$

〈표 8〉 비평균과의 차이 데이터 - 1995년을 제외한 4개년 데이터 기준

연도 산업	1994	1996	1997	1998	1999	2000	2001
전산업	+0.010	+0.021	-0.003	-0.028	-0.025	+0.008	+0.043
C	+0.066	-0.038	-0.002	-0.026	-0.032	+0.020	+0.029
D	+0.024	+0.011	-0.012	-0.024	+0.028	+0.051	+0.075
E	-0.089	+0.035	+0.021	+0.033	-0.047	-0.007	-0.045
F	+0.071	+0.018	+0.004	-0.093	-0.134	+0.015	+0.032
G	+0.008	+0.071	+0.033	-0.110	-0.059	+0.040	+0.153
H	-0.045	+0.007	-0.007	+0.046	-0.091	+0.182	+0.082
I	+0.003	+0.024	-0.005	-0.021	-0.039	-0.085	-0.029
J	+0.028	+0.027	-0.026	-0.029	-0.137	-0.049	-0.100
K	-0.040	+0.055	+0.010	-0.025	-0.146	-0.181	-0.008
M	-0.039	+0.023	+0.011	+0.006	+0.020	+0.004	+0.033
N	+0.026	+0.054	-0.020	-0.060	+0.002	-0.060	+0.059
O	-0.043	-0.020	+0.012	+0.051	-0.103	-0.085	-0.301

이와 같은 형식의 표를 구축하면 비차이값에 붙은 부호(+,-)만 보더라도 칸별로 두 조사 중 어느 쪽이 과대한지 또는 과소한지, 어떤 산업 추계에서 어떤 패턴이 존재하는지, 또 다년간 일관된 과대성 또는 과소성이 존재하는지 등을 일목요연하게 관찰할 수 있으며, 〈표 7〉과 결합하면 그 상대적 차이를 간편히 사정할 수 있다.

결국 조사표와 조사방법이 바뀐 《노동력수요동향조사》의 추계치들의 변동은 연구대상 기간 중 조사방법에 큰 변화가 없었던 《임금구조기본통계조사》를 기준으로 판단하면 새로운 조사방법의 도입 전후로 상당히 유의한 차이가 존재함을 시사하고 있다. 그러나 이러한 차이가 실제함을 확

인하더라도 《노동력수요동향조사》의 결과가 《임금구조기본통계조사》보다 더 정확함을 의미하는지는 알 수 없다. 다만 조사결과가 과거에 비하여 다르게 나오고 있음을 알 수 있을 뿐이다.

#### IV. 결어

이 연구에서는 현행 《노동력수요동향조사》의 효율성을 제고하기 위한 방편으로 새로운 조사표와 조사방법의 도입으로 인한 영향 평가 방법을 제시하였다. 또한 이 과정에서 《노동력수요동향조사》와 《임금구조기본통계조사》를 비교하여 노동부 데이터 간의 관계를 파악할 수 있는 부수 효과가 있었다.

새로운 조사표와 조사방법이 조사결과에 영향을 미쳤다면 어떤 이유에서 그렇게 되었는지 파악함이 시급하다. 앞으로 거듭될 조사들 간에 동일 모수에 대한 추정값에 큰 차이가 나는 현상이 지속됨은 지양해야 하기 때문이다.

#### 참고문헌

- 김병조·이건. 2001. “정부통계의 개선 방안 : 《노동력수요동향조사》.” 《조사연구》 2(2): 61-82.
- 김수곤·윤석천. 1998. 《2000년대 인력자원 전망과 정책과제》. 미래인력연구센터.
- 남성일. 1998. “직업전망 조사연구 사업에 대한 제언.” 미간행보고서. 노동부.
- 노동부. 2001. 《노동력수요동향조사보고서》.
- 동국대학교 자연과학연구원. 1998. 《노동력 수요동향조사 표본설계 및 조사표개발에 관한 연구》. 미간행보고서. 노동부.

- 어수봉. 1994. 《우리나라 노동통계의 주요특징》. 한국노동연구원.
- 이병직. 1995. 《한국 노동통계의 현황과 과제》. 미간행보고서. 노동부.
- 한진수·최강식. 1997. “정보통신 전문인력의 수급현황 및 전망.” 《정보통신 인적자원개발 정책보고서》. 미간행보고서. 정보통신부.
- 홍두승. 1997. “사회발전과 직업구조의 변화.” 《한국사회 50년: 사회변동과 재구조화》. 서울대학교 출판부.
- Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. *Regression Diagnostics*. Wiley.