

사업체노동력조사를 활용한 빈 일자리 수 추정에 대한 정합성 연구*

박승환
강원대학교 정보통계학과 조교수

A Study of Consistency in Estimating the Number of Vacant Jobs Using the Labor Force Survey at Establishments

Seung-Hwan Park^a

^a Department of Information Statistics, Kangwon University, South Korea

Received 18 August 2022, Revised 23 September 2022, Accepted 28 September 2022

Abstract

Purpose - The purpose of this study was to investigate consistency in estimating the number of vacant jobs using the two business labor force survey with two different time points of survey.

Design/methodology/approach - We studied the cause of the differences in estimating the number of vacant jobs between the monthly sample and the new sample in business labor force survey.

Findings - To summarize our findings, As the size of the company increases, the number of vacant jobs in the company also increases, and the probability that the number of vacant jobs in the company is zero decreases. The monthly sample was assessed to have a higher likelihood that the number of vacant jobs in the company was zero and the number of vacant jobs was considerable compared to the local sample.

Research implications or Originality - Because local survey sample companies tend to minimize the number of vacant jobs even when they reply under the same conditions, the estimation result of the number of vacant jobs in the current monthly survey differs significantly from the estimation result of the local survey. Divergent “degrees of knowledge of question items,” survey methodologies, or investigators could be the causes of the various response trends.

Keywords: Complex Sampling Design, MCMC, Multilevel Model with Overdispersion, Zero-Inflated Poisson Model

JEL Classifications: C13, C30, C83

I. 서론

고용통계는 노동의 공급측면(가구)을 파악하는 조사와 수요측면(사업체)을 파악하는 조사로 구분할 수 있다. 공급측면 고용통계로는, 가구조사를 통해 노동시장 동향을 파악하는 통계청의 경제활동인구조사 와 지역별 고용조사가 있다. 수요측면 고용통계로는 사업체 조사를 통해 고용 변동을 파악하는 고용노동부

* 이 논문은 이기재, 박승환, 박현아. (2019). 산업 소분류 고용통계생상을 위한 지역별 사업체노동력조사 추추정방법 연구 보고서의 일부 내용을 중심으로 수정, 보완 및 추가 구성하여 작성하였음.

^a First Author, Corresponding Author, E-mail: stat.shpark@kangwon.ac.kr

© 2022 The Institute of Management and Economy Research, All rights reserved.

의 사업체노동력조사와 지역별 사업체노동력조사가 있고, 행정통계로 고용보험통계가 대표적이다(이기재 외 2인, 2019).

사업체노동력조사는 경제활동인구조사와 비교할 수 있는 노동수요 측의 대표통계이다. 고용노동부에서는 2011년부터 사업체 대상의 노동통계조사를 통합하여 사업체노동력조사, 직종별 사업체노동력조사, 지역별 사업체노동력조사 등으로 실시하고 있다. 각 조사는 동일 기준시점에서 조사되지만, 표본 규모와 공표 수준이 달라서 조사가 진행됨에 따라 각 공표 수준에서 순차적으로 고용동향 관련 통계가 공표되고 있다.

현행 고용통계는 사업체노동력조사와 지역별 사업체노동력조사를 통해 작성 중이다. 사업체노동력조사는 매월 조사되며 산업중분류 수준에서 공표가 이루어지고 있으며 지역별 사업체노동력조사는 보다 세부적인 산업별 고용구조를 모니터링할 수 있도록 산업소분류별 고용 현황에 대해 공표를 하고 있다.

지역별 사업체노동력조사에서 가중치 산출 방법은 기본적으로 사업체노동력조사 결과를 활용하여 벤치마킹하고 있다. 최신 모집단(예를 들어 19년 12월 말 기준) 노동실태 현황의 산업대분류(또는 산업초대분류) × 규모 내 지역별 비중에 따라 20년 3월 말 기준 사업체노동력조사 결과의 근로자 총수를 비례배분하고, 종사자 수에 대해서 벤치마킹하는 방안을 적용하고 있다.

종사자 수를 제외한 빈 일자리 수, 입/이직자 수 등의 고용현황에 대해서는 조사대상과 조사 기준시점이 동일하지만 통계작성 수준과 표본 규모가 다르기 때문에 서로 다른 시점에서 다른 통계작성 단위로 발표하는 경우에는 두 조사 간 고용동향 통계의 불일치 문제가 발생할 수 있다. 빈 일자리수, 입/이직자 수는 단순한 종사자 수 보다 통계를 사용하는 사람들의 관심이 높은 항목이다. 이 항목들에 대한 두 조사간의 통계 불일치는 전체 조사 결과에 대한 신뢰성을 떨어뜨려 고용통계 전체의 활용에 어려움을 줄 수 있다.

이와 같은 빈 일자리 수 등의 고용통계 불일치 문제를 해결하기 위해서는 작성된 두 통계의 불일치가 발생한 원인을 찾는 것이 무엇보다 중요하다. 본 논문에서는 빈 일자리 수 추정을 중심으로 같은 시점에서 조사된 동일한 조사항목의 고용동향 관련 통계에 대해서 정합성 확보 방안에 대한 연구를 진행한다. 두 통계의 불일치가 발생한 원인을 찾기 위하여 빈 일자리 수 분포와 빈 일자리 수가 존재하는 사업체에 대한 회귀 모형을 통한 분석을 통해 빈 일자리 수 통계의 정합성 확보 방안을 연구한다.

본 논문의 구성은 다음과 같다. II장에서는 사업체노동력조사의 개요와 빈 일자리 수 추정 방법을 살펴보고 빈 일자리 수 추정 현황을 분석한다. III장에서는 빈 일자리 수 추정의 정합성 확보를 위한 회귀분석 모형 추정 방법론을 제안한다. IV장에서는 III장에서 제안한 0평창 포아송 회귀 모형과 과대산포가 존재하는 다수준 로지스틱 회귀 모형을 2018년 지역별 사업체노동력조사에 대하여 적용하고 그 결과를 해석한다. V장에서는 본 연구의 분석 내용을 정리하고 그 시사점 및 의의를 기술한다.

II. 사업체노동력조사 개요 및 빈 일자리 수 현황 분석

이기재 외 2인 (2019)를 중심으로 사업체노동력조사와 지역별 사업체노동력조사의 개요와 표본설계 요약을 통해 두 조사 간 표본설계 상 차이가 있는지 살펴본다. 두 조사의 가장 큰 표본설계 상 차이는 사업체노동력조사는 매월 약 23,000개 사업체를 조사하는 반면 지역별 사업체노동력조사는 4월과 10월 두 차례 약 200,000개 사업체를 조사한다. 지역별 사업체노동력조사의 표본은 해당월 사업체노동력조사의 표본을 포함하고 있는 것이 특징이다. 또한 두 조사에서 동일하게 사용하고 있는 종사자 수 및 빈 일자리 수 추정 방법에 대해서도 두 조사 간 고용통계 차이를 발생시킬 수 있는 부분이 있는지 살펴본다.

1. 사업체노동력조사 개요

고용노동부에서는 2011년부터 사업체 대상의 노동통계조사를 통합하여 사업체노동력조사, 직종별 사업체노동력조사, 지역별 사업체노동력조사 등으로 실시하고 있다. 경제활동인구조사가 노동 공급 측의 대표 통계라면 이에 비교할 수 있는 노동 수요 측의 대표통계는 사업체노동력조사이다.

사업체노동력조사는 기본적으로 기존 사업체고용동향조사와 사업체임금·근로시간조사를 통합한 형태

의 조사이다. 사업체노동력조사를 통해서 작성되는 주요 통계는 산업분류 및 사업체 규모별 종사자 수, 빈 일자리 수, 입.이직자 수, 근로일수, 근로시간, 임금(정액급여, 초과급여, 특별급여) 등이다.

사업체노동력조사의 조사대상은 농업, 임업 및 어업(A)과 가사 서비스업(T), 국제기구 및 기타 외국기관(U)을 제외한 전 산업의 종사자(상용, 임시·일용, 기타종사자) 1인 이상을 고용하고 있는 전체 사업체이다. 이 조사의 고용 부문은 국가 또는 지방행정기관 등 공무원 재직기관도 조사대상에 포함하고 있다.

현행 사업체노동력조사의 시도별 고용통계 산출을 위한 표본설계는 17개 시도별 산업중분류 및 사업체 규모별 안정적인 고용통계 산출을 목표로 각 시도별 표본크기를 결정하였다. 표본배분과 표본추출은 시도별 산업중분류 및 사업체 규모별 표본크기 결정 방안에 기초하였으며, 상용근로자 300인 이상의 사업체에 대해서는 전수추출하는 것을 원칙으로 하고, 300인 미만 규모의 표본층에 대해서는 시도 내 산업중분류 및 사업체 규모별(5인 미만, 5-29인, 30-299인) 표본크기를 세부 규모 층에 대하여 배분하였다. 세부 규모 층에 대한 표본배분은 사업체 규모별 사업체 수의 제곱근비례배분법을 적용하였다. 이후 세부 층에서 표본추출은 산업소분류 구분과 임시 및 일용근로자 비율에 따라 정렬한 후 계통추출법을 적용하여 추출하였다.

사업체노동력조사의 근로자 수 추정 방법은 미국 CES에서 사용되고 있는 근로자 수 추정 방법인 WLR(Weighted Link Relative) 방법을 사용하고 있다. 이 추정법은 당월과 전월 조사에 모두 응답한 표본사업체를 대상으로 고용변동률(sample link relative)을 구해 이를 전월 총 고용 규모(추정값)에 곱하여 추정하는 방식이다.

2. 지역별 사업체노동력조사 개요

지역별 사업체노동력조사는 전체 종사자(자영업자 및 무급가족종사자를 제외한 상용근로자, 임시 및 일용 근로자, 기타 근로자만을 의미하며, 이하 종사자로 표기함) 1인 이상 사업체를 대상으로 종사자 현황, 노동이동 동향 등에 대한 현황 파악을 주된 목적으로 한다. 구체적인 조사항목은 사업체의 전월 및 당월 말 종사자 수, 빈 일자리 수, 근로자의 이동(입.이직) 상황 등이다.

이 조사는 각 시군구 및 산업초대분류별 고용 관련 다양한 통계 생산이 가능하도록 표본설계되었다. 기본적으로 시군구의 지방자치단체는 지역별 전체 및 산업초대분류별 고용통계를 생산하도록 하였다.

지역별 사업체노동력조사의 조사대상은 농림어업(A), 가사서비스업(T), 국제기구 및 기타 외국기관(U) 부분을 제외한 전 산업(18개 산업대분류)에 대하여 상용근로자와 임시 및 일용 근로자, 기타 근로자 등 근로자 1인 이상을 고용하고 있는 사업체의 전체 종사자이다. 표본설계에서 활용한 표본추출틀은 2008년 12월 기준의 사업체노동실태현황결과의 근로자 1인 이상 사업체 전체 명부이다. 국가 또는 지방행정기관, 군.경찰 및 국.공립교육기관도 조사대상에 포함된다.

현행 표본설계에서 층화는 시군구 지역(229개)으로 1차 층화하고, 산업대분류(19개)별로 2차 층화해 4,351(229×19)개의 층을 부모집단(sub-population)으로 간주하여 상용근로자 수 규모를 사업체 규모의 층화변수로 선정하여 3차 층화하였다.

2018년 지역별 사업체노동력조사는 2017년 12월 말 기준의 사업체노동실태현황 결과의 사업체 명부를 표본틀로 사용하여 표본층에 대해 시군구 내 산업대분류별 사업체 규모 층에서 사업체의 종사자 순으로 정렬한 후, 각 층마다 계통추출(systematic sampling)로 표본을 추출하였다.

3. 빈 일자리 수 추정 방법

근로자 수는 매월 표본조사 결과로부터 WLR(Weighted Link Relative) 방법에 의해서 추정된 결과와 고용보험 DB 자료의 누계 총량 방식을 통해서 집계된 신설 사업체에서 발생한 고용변동의 합으로 추정하고 있다.

한국조사연구학회(2015)는 최근 신설된 사업체에서 발생하는 고용변동이 기존 사업체의 경우와 달리 고용증가가 크다는 점을 반영하여, 현행 고용보험 DB 자료에 기초한 고용 순변동(NBD) 적용방식이

아닌 고용보험 DB 자료의 누계 총량 방식으로 변경할 것을 제안한 바 있다.

현행 사업체노동력조사는 시도별 고용통계의 안정적인 생산을 위해서 각 시도별 고용통계를 산출한 후 이를 집계하는 방식으로 고용통계를 생산하고 있다. 시도별 추정법은 전월 고용 추정값에 전월 대비 해당 월의 근로자 수 변동비(ratio)를 연쇄적으로 곱하여 산출한다. 이에 따라 각 시도에서 산업중분류 및 사업체 규모 구분을 기초 추정셀로 하는 추정 방안을 적용하였다. 기초 추정셀은 각 시도 단위에서 산업중분류 × 규모(30인 미만, 30-299인, 300인 이상)를 이용하는 것을 원칙으로 하고 있다.

표본조사로부터 각 기초 추정셀의 근로자 총수 추정식은 다음과 같다.

$$\widehat{AE}_c = \widehat{AE}_p \times \frac{\sum_i (w_i \times ae_{ci})}{\sum_i (w_i \times ae_\pi)}$$

여기서, $i = \text{matched sample unit}$

$h =$ 층 첨자로 산업중분류, 규모, 임시·일용근로자 고용 여부로 구분

$w_i =$ i 번째 사업체에 대한 가중치 $= N_h/n_h$

(N_h : 층 h 의 모집단 사업체 수, n_h : 층 h 의 표본 사업체 수)

$ae_\pi = i$ 번째 사업체에 대한 전월 근로자(상용+임시·일용) 수

$ae_{ci} = i$ 번째 사업체에 대한 당월 근로자(상용+임시·일용) 수

사업체노동력조사의 기초 추정셀에서 빈 일자리 수 및 입·이직자 수에 대한 추정량은 비추정량(ratio estimator)의 형태로 다음 식에 따라 구한다.

$$\widehat{JO}_{estcell} = \frac{\sum_{i \in estcell} w_i \times JO_i}{\sum_{i \in estcell} w_i \times e_i} \times \widehat{E}_{estcell} = \sum_{i \in estcell} BMKF_i \times w_i \times JO_i$$

$$\widehat{X}_{estcell} = \frac{\sum_{i \in estcell} w_i \times x_i}{\sum_{i \in estcell} w_i \times e_i} \times \widehat{E}_{estcell} = \sum_{i \in estcell} BMKF_i \times w_i \times x_i$$

빈 일자리 수 또는 입·이직자 수의 총계 추정량에 대한 근사적인 분산 추정량은 다음과 같다.

$$\widehat{Var}(\widehat{Y}) = \sum_{h=1}^H \frac{n_h(1-f_h)}{(n_h-1)} \sum_{i=1}^{n_h} (u_{hi} - \bar{u}_h)^2$$

여기서, $N_h =$ 층 h 의 모집단 사업체 수, $n_h =$ 층 h 의 조사 사업체 수,

$y_{hi} =$ 층 h 의 i 번째 사업체의 빈 일자리 수 또는 입·이직자 수,

$$f_h = n_h/N_h, \quad \widehat{Y} = \sum_{i=1}^{n_h} w_i^f \cdot y_{hi}, \quad u_{hi} = w_{hi} \cdot y_{hi} - \widehat{Y}, \quad \bar{u}_h = \left(\sum_{i=1}^{n_h} u_{hi} \right) / n_h$$

4. 빈 일자리 정합성 분석

지역별 사업체조사와 매월 사업체 조사에 대하여 가중치를 사용한 가중 종사자 수 총계와 가중 빈 일자리 수 총계를 추정하고 95% 신뢰구간을 구한 결과를 <Table 1>에 제시하였다. 지역조사와 매월조사의 종사자 수 총계는 유사하게 나타났으나 지역조사의 빈 일자리 수 총계 추정값과 월 조사의 빈 일자리

수 총계 추정값은 차이가 크게 나타났다. 매월조사의 빈 일자리 수 총계와 빈 일자리 수가 존재하는 사업체 수 모두 지역조사의 추정값들에 비하여 상당히 크게 나타났다. 매월조사의 추정값이 지역조사 추정값의 95% 신뢰구간에 속하지 않는 것으로 보아 두 조사의 추정값은 통계적 유의한 차이가 있다고 할 수 있다고 보인다.

Table 1. Estimation of the Total Number of Employees and Vacant Jobs in Local Survey and Monthly Survey and 95% Confidence Intervals Result

항목	지역조사	95% 신뢰구간		매월조사
종사자수 총계	16,038,201	15,858,792	16,217,609	15,932,685
빈일자리수 총계	182,880	178,420	187,339	194,365
빈일자리 존재 사업체수	99,055	96,869	101,240	104,513
빈일자리 존재 사업체 비율	5.7%	5.5%	5.8%	6.0%

사업체의 특성에 가장 큰 영향을 줄 수 있는 산업대분류와 사업체 규모를 고려하여 두 조사의 빈 일자리 수 추정 결과를 비교 분석한다. <Table 2>에는 산업대분류별 지역조사와 월별조사의 빈 일자리 수 총계추정에 대한 결과가 나타나 있다. 빈 일자리 수는 산업대분류 C에서는 지역조사의 추정값이 매월조사에 비해 더 크게 나타나지만 대부분 산업대분류에서는 매월조사의 추정값이 크게 나타난다.

Table 2. Comparison of the Estimation Result of the Number of Vacant Jobs by Industrial Classification

산업대분류	지역조사		매월조사		차이 (지역 - 매월)
	빈일자리수	표준오차	빈일자리수	표준오차	
B	110.9	21.7	139.3	85.6	-28.4
C	50,512.0	929.8	44,776.0	2,587.8	5,736.0
D	45.8	12.4	29.7	11.7	16.1
E	863.4	90.0	945.1	204.4	-81.7
F	12,143.0	525.2	14,161.0	4,841.4	-2,018.0
G	25,510.0	660.4	29,271.0	2,983.4	-3,761.0
H	17,006.0	738.9	19,832.0	2,242.6	-2,826.0
I	18,843.0	603.5	27,804.0	7,297.7	-8,961.0
J	5,902.3	349.3	4,387.0	559.7	1,515.3
K	951.3	112.1	1,108.3	238.3	-157.0
L	2,925.0	272.0	2,182.7	439.9	742.3
M	8,549.1	424.8	8,069.0	999.4	480.1
N	8,221.7	508.4	6,194.2	715.0	2,027.5
O	1,029.3	139.6	1,013.8	398.0	15.6
P	3,232.6	253.6	4,374.8	925.5	-1,142.2
Q	17,837.0	1,352.7	20,174.0	2,729.1	-2,337.0
R	2,602.3	230.7	2,280.6	343.9	321.6
S	6,595.3	366.9	7,622.2	1,017.8	-1,026.9

<Table 3>에는 사업체 규모별 지역조사와 매월조사의 빈 일자리 수, 입/이직자 수 총계추정에 대한 결과가 나타나 있다. 업체 규모는 상용 근로자를 기준으로 다음과 같이 구분한다. (5인 미만, 5~9인, 10~29인, 30~99인, 100~299인, 300~999인) 빈 일자리 수는 10~29인 사업체에서는 지역조사의 추정값이 매월조사에 비해 더 크게 나타나고 대부분 다른 규모의 사업체에서는 매월조사의 추정값이 크게 나타난다. 특히 5인 미만 사업체에서 두 조사의 추정값의 차이가 크게 나타난다.

Table 3. Comparison of the Estimation of the Number of Vacant Jobs by Size of Establishment

사업체규모	지역조사		매월조사		차이 (지역 - 매월)
	빈일자리수	표준오차	총계추정	표준오차	
5인미만	53,909.0	1,171.0	65,926.0	9,308.1	-12,017.0
5-9인	28,533.0	760.7	29,040.0	2,717.0	-507.0
10~29인	39,064.0	1,471.4	35,783.0	2,217.5	3,281.0
30~99인	32,708.0	690.3	33,665.0	2,459.6	-957.0
100~299인	22,383.0	754.2	23,087.0	1,775.3	-704.0
300~999인	6,281.5	125.9	6,864.0	336.4	-582.5

지역별 조사에는 지역조사 표본과 매월조사 표본이 구분되어 있다. 두 조사의 빈 일자리 수 추정값의 차이에 대한 분석을 위하여 두 표본 간 응답 특성의 차이를 분석한다. 먼저 지역조사 내 지역 표본과 매월 표본 간 가중치 크기의 차이가 있는지 살펴본다. <Table 4>에는 지역조사의 지역 표본과 매월 표본의 가중치와 월별조사의 가중치의 분포와 기초통계량이 나타나있다. 지역조사 내 두 표본에 부여된 가중치의 평균과 분포는 두 표본에서 크게 차이가 없음을 알 수 있다. 따라서 빈 일자리 수 추정의 차이는 가중치보다는 두 표본의 빈 일자리 수 조사값의 분포 차이에서 발생함을 알 수 있다.

빈 일자리 수 유무와 빈 일자리 수 분포에 대하여 두 표본 간 응답 특성의 차이를 분석한다. 지역 표본과 매월 표본 중 빈 일자리 수가 존재하는 사업체의 수는 각각 12,523개와 2,813개이며 각 표본 내에서 빈 일자리 수가 존재하는 사업체의 비율은 지역 표본은 전체 177,610개 중 12,523개로 약 7.1%, 매월 표본은 전체 22,964개 중 2,813개로 약 12.2%로 매월 표본의 빈 일자리 수가 존재하는 사업체 비율이 지역 표본의 사업체 비율에 비하여 훨씬 크게 나타난다.

Table 4. Distribution of Weights of Local Sample Monthly Sample in Local Labor Force Survey at Establishments

가중치	지역조사	
	지역표본	매월표본
표본수	177,610	22,964
평균	8.8	8.1
표준편차	8.1	9.4
변동계수	108.1	86.2
최대값	214.2	212.8
99%	36.5	42.2
95%	24.7	25.5
90%	19.3	19.7
Q3	11.7	11.0
중위수	6.3	4.9
Q1	3.2	1.7
10%	1.8	1.0
5%	1.4	0.8
1%	0.9	0.5
최소값	0.0	0.0

빈 일자리가 있는 사업체만을 대상으로 매월/지역 표본의 빈 일자리 수 분포를 비교하였다. 가중치를 사용하지 않고 자료의 값만을 비교하여 두 표본 간의 빈 일자리 수 분포를 분석한 결과가 <Table 5>에 나타나 있다. 지역조사의 매월 표본의 빈 일자리 수의 값이 지역 표본의 값에 비하여 전반적으로 크게 나타나는 경향이 있다. 이는 매월 표본의 빈 일자리 수 평균, 중위수, Q3의 값들이 지역 표본의 값에 비하여 큰 것으로 확인할 수 있다. 빈 일자리 수 총계추정에서 지역조사의 추정값이 작게 나타나는 이유는 두 표본의 가중치 값에는 큰 차이가 없기 때문에 지역 표본의 빈 일자리 수 자료값 자체가 작기 때문이라고 해석할 여지가 있다고 판단된다.

Table 5. Distribution of the Number of Vacant Jobs of Local Sample Monthly Sample in Local Labor Force Survey at Establishments

빈일자리수	지역조사	
	지역표본	매월표본
표본수	12,523	2,813
최대값	304.0	178.0
99%	20.0	49.6
95%	7.0	20.0
90%	5.0	10.0
Q3	2.0	4.0
중위수	1.0	2.0
Q1	1.0	1.0
10%	1.0	1.0
5%	1.0	1.0
1%	1.0	1.0
최소값	1.0	1.0

III. 빈 일자리 수 분석을 위한 회귀 모형

1. 영팽창 포아송 모형

빈 일자리 수 총계추정에서 지역조사의 추정값이 작게 나타나는 현상에 대하여 다른 설명 변수의 영향을 고려한 분석을 수행한다. 빈 일자리 수와 같이 이산형 확률변수에 대하여 고려할 수 있는 회귀 모형으로는 포아송 회귀 모형, 음이항 회귀 모형, 0팽창 포아송 회귀 모형 등이 있다. 빈 일자리 수에 대하여 평균보다 분산이 큰 과대산포(overdispersion) 현상이 발생하는 점을 고려하기 위하여 0팽창 모형을 고려하였다. 이를 통해 사업체노동력조사와 지역별 사업체노동력조사의 빈 일자리 수 추정에 대한 정합성을 확보하는 방안을 찾고자 한다.

0팽창 포아송 모형은 다음과 같이 ϕ_i 의 확률로 0을 갖고 $1-\phi_i$ 확률로 0이 아닌 $g(y_i|x_i)$ 분포를 따르는 모형을 갖으며 이때 $g(\cdot)$ 은 포아송 분포를 따른다.

$$y_i \sim \begin{cases} 0 & \text{with probability } \phi_i \\ g(y_i) & \text{with probability } 1-\phi_i \end{cases}$$

설명변수 x_i 와 z_i 를 고려한 모형은 다음과 같다. 여기서 Y_i 는 이산형 확률변수이고, x_i 는 확률변수

Y_i 에 영향을 미치는 설명변수이고, z_i 는 확률변수 Y_i 가 0일 확률에 영향을 미치는 설명변수다. 0팽창 확률 모형 $\phi(\cdot)$ 은 로지스틱 함수를 가정한다.

$$P(Y_i = y_i | x_i, z_i) = \begin{cases} \phi(\gamma'z_i) + (1 - \phi(\gamma'z_i))g(0|x_i) & \text{if } y_i = 0 \\ (1 - \phi(\gamma'z_i))g(y_i|x_i) & \text{if } y_i > 0 \end{cases}$$

모수 추정에는 최대우도추정량(MLE)으로 계산할 수 있으며 SAS의 PROC COUNTREG 프로시저를 사용하여 모수 추정을 수행하였다.

2. 과대산포가 존재하는 다수준 로지스틱 회귀 모형

빈 일자리 수에 0의 발생이 과도하다는 현상을 추가로 분석하기 위하여 빈 일자리 수 유무에 대한 로지스틱 회귀분석을 수행하였다. 사업체 조사자료는 구조적으로 사업체의 산업분류에 대하여 다층구조(multilevel)를 갖고 있다. 따라서 지역별 사업체조사의 빈 일자리 수에 대한 다수준 회귀 모형을 통한 분석을 수행한다.

빈 일자리 수 존재 여부를 각 사업체별로 얻을 수 있으나 이를 그룹화하여 이항분포를 가정한 모형을 고려한다. 여기서 그룹은 사업체규모 \times 표본구분으로 생성하였다. 따라서 각 그룹의 사업체들은 같은 사업체 규모, 표본 구분 값을 갖게 되며 추가적인 보조 정보를 통한 사업체의 구분은 불필요하다. 그룹 i 에 대하여 다음의 이항분포 모형을 생각할 수 있으며 빈 일자리 존재 여부에 대한 성공확률 p_i 에 대한 로지스틱 회귀 모형을 고려하는 경우 모형은 다음과 같다.

$$y_i \sim \text{binomial}(n_i, p_i) \\ \text{logit}(p_i) = X_i^T \beta$$

이러한 로지스틱 회귀 모형에서 y_i 의 분산이 평균보다 큰 과대산포 현상을 반영하기 위하여 Browne et al. (2005)에서 다음의 다수준 로지스틱 회귀 모형을 고려하였다.

$$y_{ij} \sim \text{binomial}(n_{ij}, p_{ij}), \\ \text{logit}(p_{ij}) = X_{ij}^T \beta + u_{ij}, \\ u_{ij} \sim N(0, \sigma_u^2)$$

이 모형에서는 과대산포를 반영하기 위하여 2 수준 모형을 고려하였는데 2 수준의 개체는 각각 한 개의 1 수준 개체를 갖고 있다고 가정한다. 다음은 산업분류를 추가로 고려하여 과대산포를 반영한 3 수준 로지스틱 회귀 모형이다.

$$y_{ijk} \sim \text{binomial}(n_{ijk}, p_{ijk}), \\ \text{logit}(p_{ijk}) = \beta_0 + X_{ijk}^T \beta + v_k + u_{jk}, \\ v_k \sim N(0, \sigma_v^2) \\ u_{jk} \sim N(0, \sigma_u^2) \\ j = 1, \dots, n_k, \quad k = 1, \dots, K.$$

여기서 i 는 그룹을 뜻하며 j 는 과대산포를 위한 개체, k 는 산업분류를 뜻한다. v_k 는 산업분류로 인한 랜덤효과(random effect)를 뜻하고 n_k 는 산업분류 k 에 속한 그룹의 개수이다.

모수 추정에는 MCMC알고리즘을 통하여 이루어졌으며 WinBUGS 1.4(Spiegelhalter et al., 2003)를

통하여 구현되었다. MCMC알고리즘을 구현할 때 v_k 와 u_{jk} 를 직접 생성하지 않고 재매개화(reparametrization)를 통하여 $u_{jk}^* = X_{jk}^T \beta + u_{jk} + v_k^*$, $v_k^* = \beta_0 + v_k$ 를 생성하도록 한다. 사전분포로 β 에 대하여 무정보 사전분포인 flat prior를 주고 σ_v^2 과 σ_u^2 에 대해서는 인버스 감마 분포를 준다. 변인은 500개로 하였고 10,000개의 반복을 통한 MCMC알고리즘을 구현하였다.

IV. 실증분석 결과

2018년 4월 지역별 사업체 조사자료를 통하여 빈 일자리 수에 영향을 미치는 보조변수와의 관계를 살펴보았다. 설명변수로 사용한 사업체의 주요 특성 변수를 재정의 하였다. 각 변수에 대한 정의와 특성은 <Table 6>에 나타나 있다.

Table 6. Variables Used in Models

변수	정의	특성
사업체규모	종사자 수 기준의 사업체 규모	1=1~4인, 2=5~9인, 3=10~29인, 4=30~100, 5=100~299인, 6=300~999인, 7=1000인 이상
권역	권역 구분	1=서울, 2=인천, 경기, 강원, 3=대전, 세종, 충북, 충남, 4=광주, 전북, 전남, 제주, 5=부산, 대구, 울산, 경북, 경남, 6=권역통합
산업초대분류	산업대분류 범주화	1=B, C, 2=F, 3=G, I, 4=E, L, M, N, O, P, Q, R, S, 5=D, H, J, K
산업중분류	산업중분류	
표본구분	지역 표본 or 매월 표본	지역, 매월

1. 영팽창 포아송 모형 추정 결과

다음은 2018년 4월 지역별 사업체노동력조사에 대하여 0팽창 포아송 모형(Zero-Inflated Poisson Model)을 적용하여 분석한 결과이다. 빈 일자리 수에 대한 0팽창 포아송 모형 추정 결과는 <Table 7>과 <Table 8>에 포아송 회귀 모형과 0팽창 확률 모형에 대하여 각각 나타나 있다. 먼저 포아송 회귀분석 결과를 보면 권역, 산업초대분류, 사업체 규모, 매월 표본 여부가 유의한 변수로 나타난다. 매월 표본 여부와 산업초대분류의 교호작용, 매월 표본 여부와 사업체 규모의 교호작용에 대하여는 유의한 일부 경우만 결과로 제시하였다.

산업대분류 D, H, J, K에서 다른 산업대분류에 비하여 빈 일자리 수가 가장 크게 나타나며 다음은 산업대분류 F이고 나머지 산업대분류에서는 유사한 빈 일자리 수 분포를 갖는다. 다음으로 사업체 규모도 모든 범주에서 유의하게 나타나며 사업체 규모가 커질수록 사업체 내의 빈 일자리 수도 늘어나는 효과를 보인다. 매월 표본은 지역 표본에 비하여 사업체의 다른 조건이 같은 경우 빈 일자리 수는 작게 나타나는 효과를 보인다.

매월 표본 여부와 산업대분류의 교호작용을 보면 산업초대분류 1, 3, 4에서는 지역 표본에 비하여 빈 일자리 수가 크게 나타나며 산업초대분류 2에서는 지역 표본에 비하여 빈 일자리 수가 작게 나타난다. 사업체 규모와의 교호작용은 사업체 규모 5~9인에서만 유의하게 나타나며 지역 표본에 비하여 빈 일자리 수가 매월 표본에서 크게 나타난다.

Table 7. Zero-Inflated Poisson Regression: the Result of Parameter Estimates and Standard Errors for Poisson Regression

설명변수	회귀계수	S.E	유의확률
상수항	2.765	0.046	<.0001
권역1	0.503	0.039	<.0001
권역2	0.394	0.038	<.0001
권역3	0.391	0.039	<.0001
권역4	0.405	0.041	<.0001
권역5	0.270	0.039	<.0001
산업초대분류1	-0.908	0.018	<.0001
산업초대분류2	-0.258	0.024	<.0001
산업초대분류3	-0.995	0.025	<.0001
산업초대분류4	-0.897	0.019	<.0001
사업체규모1	-2.718	0.038	<.0001
사업체규모2	-2.355	0.036	<.0001
사업체규모3	-1.963	0.031	<.0001
사업체규모4	-1.367	0.028	<.0001
사업체규모5	-0.539	0.028	<.0001
매월표본	-0.303	0.036	<.0001
매월표본*산업초대분류1	0.412	0.031	<.0001
매월표본*산업초대분류2	-0.330	0.071	<.0001
매월표본*산업초대분류3	0.184	0.045	<.0001
매월표본*산업초대분류4	0.516	0.030	<.0001
매월표본*사업체규모2	1.645	0.045	<.0001

산업대분류 D, H, J, K에서 다른 산업대분류에 비하여 빈 일자리 수가 0일 확률이 가장 크게 나타나며 다음은 산업대분류 F와 산업대분류 E, L, M, N O, P Q, R, S 순으로 빈 일자리가 0일 확률이 크다. 산업초대분류 1과 3에서는 빈 일자리 수가 0일 확률이 가장 작으며 두 산업초대분류의 효과는 비슷하게 나타난다. 다음으로 사업체 규모가 가장 작은 경우 빈 일자리가 0일 확률이 가장 크며 사업체 규모가 증가할수록 빈 일자리가 0일 확률이 점점 작아진다. 매월 표본은 지역 표본에 비하여 사업체의 다른 조건이 같은 경우 빈 일자리가 0일 확률이 더 크게 나타남을 알 수 있다.

교호작용을 보면 산업초대분류 1과 3에서는 매월 표본에서의 빈 일자리 수가 0일 확률이 지역 표본에 비하여 크게 나타나며 사업체 규모와의 교호작용에서는 전반적으로 지역 표본에 비하여 빈 일자리 수가 0일 확률이 작게 나타남을 알 수 있다.

Table 8. Zero-Inflated Poisson Regression: the Result of Parameter Estimates and Standard Errors for the Inflated Model

설명변수	회귀계수	S.E	유의확률
상수항	1.434	0.125	<.0001
권역1	0.002	0.068	0.9755
권역2	-0.189	0.066	0.0041
권역3	-0.241	0.068	0.0004
권역4	0.223	0.070	0.0014
권역5	0.273	0.067	<.0001
산업초대분류1	-1.197	0.040	<.0001
산업초대분류2	-0.378	0.046	<.0001
산업초대분류3	-0.975	0.042	<.0001
산업초대분류4	-0.277	0.040	<.0001
사업체규모1	1.971	0.108	<.0001

사업체규모2	1.387	0.108	<.0001
사업체규모3	0.855	0.107	<.0001
사업체규모4	0.372	0.106	0.0005
사업체규모5	-0.140	0.110	0.2022
매월표본	0.375	0.134	0.0052
매월표본*산업초대분류1	0.455	0.079	<.0001
매월표본*산업초대분류3	0.236	0.092	0.0106
매월표본*사업체규모1	-0.891	0.157	<.0001
매월표본*사업체규모2	-0.419	0.135	0.0019
매월표본*사업체규모3	-0.616	0.132	<.0001
매월표본*사업체규모4	-0.487	0.128	0.0001

2. 과대산포가 존재하는 다수준 로지스틱 회귀 모형 추정 결과

사업체를 그룹화하기 위해서 6개 범주로 이루어진 권역 변수를 서울 / 경기, 강원, 충청 / 경상, 전라의 3개 범주로 변환하였고 사업체 규모도 30인 이하 소규모와 30인 초과 대규모 사업체로 변경하였다. 생성된 그룹은 총 12개로 권역(3) × 사업체규모(2) × 표본구분(2) 이루어져 있다.

MCMC를 통한 모수 추정 결과가 <Table 9>에 제시되어 있다. MCMC 추정치는 각 모수의 사후분포의 평균으로 계산되었으며 95% 신뢰구간을 함께 제시하였다. 경기, 강원, 충청 과 서울 지역의 사업체는 경상, 전라 지역의 사업체에 비하여 빈 일자리 수가 존재할 확률이 높게 나타났다. 대규모 사업체는 소규모 사업체에 비하여 빈 일자리 수가 존재할 확률이 높으며 매월 표본 사업체는 지역 표본 사업체에 비하여 빈 일자리 수가 존재할 확률이 높게 나타났다. 과대산포 분산 σ_u^2 에 대한 추정치는 0.167로 95% 신뢰구간을 보았을 때 통계적으로 유의하게 0보다 크다고 할 수 있으며 이는 빈 일자리 수 존재 여부 변수에 과대산포가 존재함을 뜻한다.

Table 9. The Result of Parameter Estimates and 95% Credible Intervals of MCMC

설명변수	MCMC 추정치	2.5%	97.5%
상수항	-1.994	-2.408	-1.582
서울	0.397	0.227	0.567
경기,강원,충청	0.559	0.398	0.721
소규모	-1.500	-1.636	-1.360
매월표본	0.134	-0.001	0.268
과대산포 분산 (σ_u^2)	0.167	0.124	0.223
산업대분류 분산 (σ_v^2)	0.636	0.292	1.316

12개 각 그룹별 빈 일자리 존재 확률과 VPC(Variance Partitioning Coefficient)를 <Table 13>에 제시하였다. VPC는 다수준 모형에서 전체 분산을 각 수준의 분산으로 나누어 각 수준의 변동이 전체 변동 중 차지하는 비율을 뜻한다. 로지스틱 모형에서 VPC는 선형모형에서의 VPC와 다르게 단순한 분산의 비로 계산되지는 않는다. 본 연구에서 VPC는 Browne et al. (2005)의 Method A를 사용하였으며 VPC 계산의 자세한 내용은 Browne et al. (2005)에서 확인할 수 있다.

빈 일자리는 경기, 강원, 충청의 대규모 사업체 매월 표본에서 존재할 확률이 가장 높으며 이 그룹의 변동성도 다른 그룹에 비하여 가장 크게 나타난다. 특히 빈 일자리 존재 확률에 대한 변동성은 소규모 사업체에 비해 대규모 사업체에서 지역 구분에 무관하게 높게 나타난다. 따라서 추후 조사 수행에 있어 대규모 사업체에서의 조사에 대한 콜백 등을 활용한 자료 수집의 정합성 확보를 위한 노력이 필요하다.

Table 10. The Predicted Probability of Existence of Vacant Jobs in Establishment and VPC Estimates for the 12 Types of Group in Model

권역	사업체규모	표본구분	빈일자리 존재 확률	VPC
경기,강원,충청	대규모	매월	0.214	9.422
경기,강원,충청	대규모	지역	0.192	8.781
경기,강원,충청	소규모	매월	0.057	3.291
경기,강원,충청	소규모	지역	0.050	2.934
전라,경상	대규모	매월	0.135	6.777
전라,경상	대규모	지역	0.120	6.183
전라,경상	소규모	매월	0.034	2.011
전라,경상	소규모	지역	0.029	1.779
서울	대규모	매월	0.188	8.645
서울	대규모	지역	0.168	8.003
서울	소규모	매월	0.049	2.863
서울	소규모	지역	0.043	2.545

V. 결론

본 연구에서는 매월 조사되는 사업체노동력조사와 반기마다 조사되는 지역별 사업체노동력조사의 빈 일자리 수 추정을 중심으로 한 고용현황 추정의 정합성 문제를 다루었다. 빈 일자리 수 총계추정에 있어 두 조사 간의 차이의 원인은 가중치가 아닌 지역 표본과 매월 표본의 빈 일자리 수 분포에 있음을 파악하고 특히 사업체 규모에 영향을 많이 받는 것을 확인했다.

빈 일자리 수에 대한 실증 분석은 ZIP 모형 분석을 통하여 여러 변수를 동시에 고려하여 이루어졌다. 빈 일자리 수에 영향을 미치는 설명변수로는 지역, 산업초대분류, 사업체규모, 매월 표본 여부 등이 있다. 특히 사업체 규모가 가장 중요한 변수로 분석되었고, 매월조사 사업체 여부도 중요한 영향을 미치고 있다. 사업체 규모가 커질수록 사업체 내의 빈 일자리 수도 늘어나고, 사업체 내 빈 일자리 수가 0일 가능성은 줄어든다. 매월 표본은 지역 표본에 비하여 사업체 내 빈 일자리 수가 0일 가능성이 높고 빈 일자리 수는 작게 나타는 것으로 분석되었다. 결론적으로 현재 매월조사의 빈 일자리 수 추정 결과가 지역조사의 추정 결과와 차이가 큰 것은 지역조사 표본 사업체는 같은 조건인 경우에 빈 일자리가 없다고 응답하는 경향이 높고 응답을 하더라도 빈 일자리 수를 축소해서 응답하는 경향이 있기 때문인 것으로 분석되었다. 이와 같이 응답 경향이 다르게 나타나는 이유로는 '질문항목에 대한 이해 정도'의 차이, 조사 방법이나 조사원의 차이 등이 있을 것이다.

과대산포가 존재하는 다수준 로지스틱 모형을 통하여 빈 일자리 수 존재 여부에 대한 심층 분석을 추가로 수행하였다. 앞선 분석과 유사하게 대규모 사업체는 소규모 사업체에 비하여 빈 일자리 수가 존재할 확률이 높으며 매월 표본 사업체는 지역 표본 사업체에 비하여 빈 일자리 수가 존재할 확률이 높게 나타났다. 또한 VPC 분석을 통하여 경기, 강원, 충청의 대규모 사업체 매월 표본에서 빈 일자리가 존재할 확률이 가장 높으며 이 그룹의 변동성도 다른 그룹에 비하여 가장 크게 나타남을 알 수 있었다. 추후 조사 수행에 있어 대규모 사업체에서의 조사에 대한 콜백 등을 활용한 자료 수집의 정합성 확보를 위한 노력이 필요하다는 점을 파악하였다.

본 연구에서는 정합성의 원인을 주로 파악하였다면 추후 연구에 있어서는 지역 표본을 기준으로 매월 표본에 대한 오차 보정 방법으로 결측값 대체법을 활용한 대규모 대체법을 고민해 볼 필요가 있다. 가중치를 개선하기보다는 조사 수행 과정에서 발생하는 오차를 보정하기 위한 대규모 대체법이 필요할 것이다.

References

- 이기재, 박승환, 박현아 (2019), "2019년 산업 소분류 고용통계생산을 위한 지역별 사업체노동력조사 추정방법연구 보고서", 고용노동부.
- 한국조사연구학회 (2015), "사업체노동력조사 추정법 개선방안 연구", 고용노동부.
- Browne, W. J., S. V. Subramanian, K. Jones, and H. Goldstein (2005), "Variance Partitioning in Multilevel Logistic Models that Exhibit Overdispersion", *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168, 599-613.
- Spiegelhalter, D. J., A. Thomas, N. Best and D. Lunn, (2003), "WinBUGS Version 1.4 User Manual", *MRC Biostatistics Unit, Cambridge*.