

본 연구는 통계청의 학술연구
용역에 의해 수행되었음.

가구부문 연동모형 및 추정방법 개선 연구

2011. 11.

한국통계학회

본 연구는 통계청의 학술연구
용역에 의해 수행되었음.

가구부문 연동모형 및 추정방법 개선 연구

2011. 11

연구수행기관 한국통계학회

책임 연구원 : 박민규 (고려대학교 통계학과 교수)

공동 연구원 : 김영원 (숙명여자대학교 통계학과 교수)

제 출 문

통계청장 귀하

본 보고서를 “가구부문 연동모형 및 추정방법 개선
연구 용역사업”의 최종보고서로 제출합니다.

2011. 11.

수행기관 : 한국통계학회

연구책임자 : 박 민 규 (고려대학교 통계학과 교수)

공동연구원 : 김 영 원 (숙명여자대학교 통계학과 교수)

연구보조원 : 김 관 수 (고려대학교 통계학과 박사수료)

김 미 정 (고려대학교 통계학과 석사과정)

곽 은 선 (숙명여자대학교 통계학과 박사과정)

한 혜 은 (숙명여자대학교 통계학과 박사과정)

■ 차례 ■

I. 연구개요	1
1. 연구 배경 및 필요성	1
2. 연구 내용	2
II. 연동 모형 관련 기존 연구결과 검토	5
1. 기존 통계청 연동모형 연구 결과 요약	5
1) 경제활동 인구조사	7
2) 도시간계 조사	8
3) 연동시스템 시험조사 연구 결과 요약 및 한계점	9
4) 현행 1/36 연동방안	14
2. 외국 연동 교체 방안 비교 연구	15
III. 연동표본추출을 위한 표본설계	19
1. 기존 표본설계의 타당성 검증	19
1) 응답 패턴 분석	19
2) 집락 효과 분석	33
3) 조사구 정렬 변수 분석	46
4) 업무량 파악을 위한 FGI 결과	53
2. 연동모형 개선안	56
1) 층화 및 정렬	56
2) 표본 크기	70
3) 연동 모형 개선을 위한 제언	88
IV. 추정량 및 분산추정량 산출방안 연구	94
1. 추정량의 종류	95
2. 현행 사후층화추정량 분석	101
3. 최량선형불편추정량 사용을 위한 시계열 모형 적합	121
4. 분산추정식	130

1) 개론	130
2) 층화단순임의추출 하에서의 분산추정	131
3) 층화집락추출 하에서의 분산추정	137
4) 경제활동인구조사 2010년 10월 자료를 이용한 분산추정량 비교	144
5) 모의실험	152
6) 2상 추출(Two-phase Sampling)하에서의 분산추정	153
 V. 결론	 156
 참고문헌	 158
 부록	 159

■ 표 차례 ■

〈표 2-1〉 경제활동 연동모형 시험조사 규모 및 지역 (단위: 가구)	6
〈표 2-2〉 도시가계 연동모형 시험조사 규모 및 지역 (단위: 가구)	6
〈표 2-3〉 경찰 인구조사 조사지역 및 표본구성	7
〈표 2-4〉 4-8-4연동시스템의 응답횟수별 응답률	7
〈표 2-5〉 6-6-6연동시스템의 응답횟수별 응답률	7
〈표 2-6〉 18개월 연속조사 연동모형 : 응답횟수별 가계부 회수율	8
〈표 2-7〉 경상 도시가계조사의 가계부 회수율	8
〈표 2-8〉 경제활동인구조사 응답률	9
〈표 2-9〉 가계조사 가계부 회수율	9
〈표 2-10〉 4-8-4모형에서 응답횟수별 편향지수(TIS)	12
〈표 2-11〉 6-6-6모형에서 응답횟수별 편향지수(TIS)	13
〈표 3-1〉 〈경찰〉 층별 응답 현황	20
〈표 3-2〉 〈경찰〉 응답 가구 및 가구원 현황	21
〈표 3-3〉 〈경찰〉 가구 이탈/진입 현황	22
〈표 3-4〉 〈경찰〉 2010년(1월 기준) 층 및 월별 응답 가구수 (단위: 가구)	23
〈표 3-5〉 〈경찰〉 2010년 층 및 응답횟수별 가구수	24
〈표 3-6〉 〈경찰〉 2010년 층 및 월별 응답 가구원수 (단위: 명)	26
〈표 3-7〉 〈경찰〉 2010년 연령 및 월별 응답 가구원수 (단위: 명)	27
〈표 3-8〉 〈경찰〉 2010년 응답횟수 및 연령별 응답자수	28
〈표 3-9〉 〈가계〉 층별 응답 현황	29
〈표 3-10〉 〈가계〉 응답 가구 현황	30
〈표 3-11〉 〈가계〉 가구 이탈/진입 현황	30
〈표 3-12〉 〈가계〉 2010년(1월 기준) 층 및 월별 응답 가구수 (단위: 가구)	31
〈표 3-13〉 〈가계〉 2010년 층 및 응답횟수별 가구수	32
〈표 3-14〉 〈경찰〉 층 및 월별 경찰인구의 집락효과	35
〈표 3-15〉 〈경찰〉 층 및 월별 취업인구의 집락효과	36
〈표 3-16〉 〈경찰〉 층 및 월별 실업인구의 집락효과	37
〈표 3-17〉 〈경찰〉 층 및 월별 비경찰인구의 집락효과	38

〈표 3-18〉 〈경찰〉 총 및 월별 취업시간(1주)의 집락효과	39
〈표 3-19〉 〈가계〉 총 및 월별 소득의 집락효과	41
〈표 3-20〉 〈가계〉 총 및 월별 경상소득의 집락효과	42
〈표 3-21〉 〈가계〉 총 및 월별 근로소득의 집락효과	43
〈표 3-22〉 〈가계〉 총 및 월별 가계지출의 집락효과	44
〈표 3-23〉 〈가계〉 총 및 월별 소비지출의 집락효과	45
〈표 3-24〉 정렬기준변수와 실업률, 경찰비율과의 상관계수	47
〈표 3-25〉 거쳐유형과 실업률과 경찰비율과의 분산분석	48
〈표 3-26〉 정렬기준변수와 가계자료 주요변수와의 상관계수	50
〈표 3-27〉 거쳐유형과 가계자료주요변수와의 분산분석	52
〈표 3-28〉 30~40대 비율과 실업률, 경찰비율과의 상관계수	57
〈표 3-29〉 30대 이상 대졸자비율과 실업률, 경찰비율과의 상관계수	58
〈표 3-30〉 4개 지역에서 실업률, 경찰비율과 거쳐유형을 제외한 정렬기준변수와의 상관계수	59
〈표 3-31〉 4개 지역에서 거쳐유형을 제외한 정렬기준변수와 가계자료 주요변수와의 상관계수	60
〈표 3-32〉 25개 총별 실업률과 거쳐유형을 제외한 정렬기준변수와의 회귀분석	62
〈표 3-33〉 총별 분류기준변수	70
〈표 3-34〉 경제활동인구조사의 총별 조사구당 최적 표본 가구수	72
〈표 3-35〉 〈경찰〉 실업인구의 총별 조사구당 최적 표본 가구수	74
〈표 3-36〉 가계동향조사의 총별 조사구당 최적 표본 가구수	75
〈표 3-37〉 〈가계〉 소비지출의 총별 조사구당 최적 표본 가구수	77
〈표 3-38〉 〈경찰〉 실업인구 총계의 시도 및 월별 상대표준오차	79
〈표 3-39〉 〈경찰〉 실업인구 총계의 총 및 월별 상대표준오차	80
〈표 3-40〉 〈가계〉 소비지출 평균의 시도 및 월별 상대표준오차	81
〈표 3-41〉 〈가계〉 소비지출 평균의 총 및 월별 상대표준오차	82
〈표 3-42〉 〈경찰〉 목표 CV별 표본 가구수	83
〈표 3-43〉 〈경찰〉 시도별 적정 표본가구수와 조사구수 1	84
〈표 3-44〉 〈가계〉 목표 CV별 표본가구수	85
〈표 3-45〉 〈가계〉 시도별 적정 표본가구수와 조사구수 1	86
〈표 3-46〉 〈가계〉 시도별 적정 표본가구수와 조사구수 2	87

〈표 3-47〉 〈경찰〉 시도별 적정 표본가구수와 조사구수 2	89
〈표 3-48〉 연동표본 교체 방안(1안)	91
〈표 3-49〉 연동표본 교체 방안(2안)	92
〈표 3-50〉 연동표본 교체 방안(3안)	93
〈표 4-1〉 연동그룹별 실업률 추정량 분석	111
〈표 4-2〉 실업률 자료(∇y_t)의 백색잡음 검정	123
〈표 4-3〉 세 가지 모형에서의 AIC와 SBC값 (ML추정방법)	124
〈표 4-4〉 상수항이 없는 ∇y_t 의 모형에서 모수 추정치와 적합도 통계치	124
〈표 4-5〉 ARI(4,1)에 대한 포트만토 검정치	125
〈표 4-6〉 JK 반복 가중치 비교	145
〈표 4-7〉 가중치가 다른 8,903명의 w100_a1과 w100_c1의 교차표	147
〈표 4-8〉 관심변수의 총계에 대한 표준오차	149
〈표 4-9〉 층별 JK 분산추정방안 A의 표준오차	150
〈표 4-10〉 층별 JK 분산추정방안 C의 표준오차	151
〈표 4-11〉 층별 관심변수의 총계 추정	152
〈표 4-12〉 총계 추정식 및 분산 추정식의 Monte Carlo Properties(성질)	153

■ 그림 차례 ■

[그림 2-1] 경찰 인구조사 조사지역 및 표본구성	11
[그림 3-1] 실업률-서울	65
[그림 3-2] 실업률-광역시	66
[그림 3-3] 실업률-동부	66
[그림 3-4] 실업률-읍면부	66
[그림 3-5] 경상소득-서울	68
[그림 3-6] 경상소득-광역시	68
[그림 3-7] 경상소득-동부	68
[그림 3-8] 경상소득-읍면부	69
[그림 3-9] 경제활동인구조사의 변수별 최적 표본 가구수 분포(5:1)	73
[그림 3-10] 가계동향조사의 변수별 최적 표본 가구수 분포(5:1)	76
[그림 4-1] 연동그룹별 추정량분석	103
[그림 4-2] 응답횟수별 실업률 변동 추이	106
[그림 4-3] 실업률 자료(y_t)의 시 도표	122
[그림 4-4] 실업률 자료(∇y_t)의 시 도표	122
[그림 4-5] 실업률 자료(∇y_t)의 ACF	123
[그림 4-6] 실업률 자료(∇y_t)의 PACF	123
[그림 4-7] 실업률 자료의 시 도표 - 예측값 및 신뢰구간	125
[그림 4-8] 각 연동그룹 추정량들의 ARI(4,1)모형 적합 결과	126
[그림 4-9] 모집단 구조	131
[그림 4-10] w100_a1과 w100_c1의 산점도 Plot	146
[그림 4-11] 가중치가 다른 8,903명의 w100_a1과 w100_c1의 BUBBLE Plot	146

I. 연구개요

1. 연구 배경 및 필요성

2005년 이후 도입되어 가구부문 경상조사(경제활동인구조사, 가계동향조사, 집세조사, 비농가양곡소비량조사)를 위한 조사기법으로 통계청에서는 일정한 주기로 표본의 일부가 교체되는 연동표본(rotation sample) 조사기법을 사용하고 있다. 연동표본 조사기법이 사용되는 이유로는 적절한 대체 표본 추출 방법이 사용된 경우, 표본의 시의성을 갖춘 모집단 대표성 보장, 전체 표본 개편이 아닌 일부 표본 대체로 인한 각 시점 별 통계량들의 시계열 유지, 그리고 독립된 표본들을 바탕으로 한 연속조사에 비하여 상대적으로 적은 조사 및 표본관리 비용 등을 고려할 수 있다. 언급된 연동표본조사 기법의 장점을 극대화하고 또한 패널조건화(panel conditioning)로 인한 추정량의 편향(bias) 발생과 같은 연동표본의 문제점을 최소화하기 위해서는 표본 대체를 위한 표본설계, 연동표본 모형, 그리고 이를 바탕으로 한 적절한 추정량과 분산 추정량에 대한 연구가 요구된다.

이러한 배경 하에서 2010년 실시된 인구주택총조사를 바탕으로 새로운 표본추출틀(sampling frame)이 구성되고 이를 이용한 4차 연동표본 추출이 요구되는 시기에 기존의 연동표본 조사방법론에 대한 전반적인 검토 및 학문적인 연구를 통한 개선방안의 도출이 이루어지는 것이 바람직할 것이다. 따라서 본 연구에서는 미국, 캐나다 등과 같은 선진국들의 연동표본 조사방법에 대한 검토와 함께, 가구부문 경상조사를 위한 연동표본 조사방법론에 대한 개선 방안 도출을 위하여 크게 표본설계, 연동모형 그리고 추정량 및 분산추정과 같은 세 가지 부문에 대한 연구가 중점적으로 이

투어졌다.

2. 연구 내용

가구부문 정상조사를 위한 연동표본 조사기법의 개선을 위하여 본 연구에서는 크게 연동표본 모형 개선, 연동표본 추출을 위한 표본설계 효율성 제고 방안, 그리고 통계적으로 타당한 추정량 및 분산 추정량 개발과 관련된 연구를 수행하였다. 각 주제별 세부 연구 내용은 아래와 같다.

1) 연동모형 관련 기존 연구결과 검토

연동모형 결정에 있어 중요한 요소는 교체방법, 교체주기 그리고 교체규모이다. 현행 경제활동인구조사의 경우, 전체 표본은 9개의 그룹으로 구성되어 있으며 연동 대상 표본 조사구가 완전히 다른 조사구로 대체되는데 약 4개월의 시간이 소요된다. 결과적으로 현행 연동모형에서는 매월 전체 가구표본의 1/36이 교체되고 있다.

경제활동인구조사를 위해 현재 우리나라에서 사용되고 있는 교체방법 이외에 선진국에서 사용되고 있는 방법으로는 미국의 Census Bureau에서 시행하는 CPS(Current Population Survey)를 위한 4-8-4방법과 이의 변형인 3-9-3방법 그리고 캐나다의 LFS(Labour Force Survey)를 위한 6달 연속조사방법 등이 있다. 여기서 4-8-4(3-9-3)란 한 응답가구가 처음 4(3)달 동안 연속적으로 표본에 포함된 후 다음 8(9)달간 표본에서 제외되었다가 그 다음 4(3)달 동안 다시 표본에 속한 후 완전히 표본에서 제외되는 표본조사 방식을 의미한다. 따라서 4-8-4 방법의 경우, 월간 표본 중복률은 75%이며 3-9-3 방법의 경우 67%의 월간 표본 중복률을 나타내어 우리나라의 97%의 월간 표본 중복률보다 낮게 나타난다.

본 연구에서는 먼저 통계청에서 수행한 기존의 연동 모형에 대한 연구결과들을 정리하는 동시에 해외 연동 모형들을 분석하였다. 해외 사례들의 비교를 위해서는 추정량, 연동 교체주기, 교체규모 등에 대한 사항을 중심으로 분석하였다.

2) 연동표본추출을 위한 표본설계

경제활동인구조사의 연동표본추출 표본설계 효율성 제고를 위해 본 연구에서는 우선 현행 조사가 얼마나 안정적으로 수행되고 있는지 확인해 보기 위해 가구 및

가구원 기준으로 응답패턴을 분석하였다. 경제활동인구조사나 가계동향조사와 같은 연속조사에서는 안정적인 통계 작성을 위해 표본에 포함된 조사단위들로부터 누락 없이 지속적으로 응답을 받아내는 것이 매우 중요하다는 점을 고려해 이런 분석을 수행하는 것이다. 특히 유동성이 강한 청년층의 경우 다른 연령대에 비해 응답패턴에 어떤 차이가 있는지를 확인해 보는 것 자체가 의미가 있을 것이다.

아울러 통계청의 가구부문 표본설계에서 적용되는 층화 집락추출의 경우 집락에서 추출하는 가구 규모를 어느 정도 수준으로 결정할 것인지에 따라 전체 표본크기나 조사비용 및 인력 운영 등에 큰 영향을 주게 된다. 따라서 본 연구에서는 향후 표본설계에서 효율성 제고를 위한 시사점을 찾는 것을 목적으로 현행 조사에서 조사구를 집락으로 사용함에 따른 집락효과(cluster effect)를 다양한 관점에서 분석하였다. 또한 가구부문 표본설계에서 지역 구분에 따른 층화 후 층 내 조사구 계통추출의 효율성 제고를 위한 내재적 층화(implicit stratification)를 위해 사용하는 현행 조사구 정렬방식의 적절성을 평가하는 동시에 향후 개선 방안을 모색하기 위해, 2005년 인구주택총조사 조사구 정보와 현행 경제활동인구조사의 표본 조사구의 특성을 연계해 경제활동인구조사 및 가계동향조사에서의 관심 변수 관점에서 실제 현행 표본설계에서 사용하는 조사구 정렬방법의 타당성을 검증하였다. 아울러 그 동안 연구자들의 경험을 토대로 추가로 활용 가능한 정렬 기준 변수를 제안하고 이 새로운 변수의 효율성을 검토하였다. 한편 현재 현장 조사 직원들이 느끼는 업무량과 향후 연동 모형을 변경하는 경우 현장 수용 가능성을 점검해 보기 위해 조사 직원들의 업무량 파악에 초점을 맞추어 경제활동인구조사를 담당하는 지방 조사 직원들을 대상으로 한 표적집단면접(FGI)을 수행하였다.

마지막으로 향후 가구부문 표본설계 개선과 관련해 본 연구에서 수행한 현행 표본설계에 대한 타당성 검증 결과를 토대로 층 내 정렬 방법 개선과 새로운 연동모형 구현 방안 등에 대해 제안하고자 한다. 본 연구에서 제안된 내용은 향후 경제활동인구조사나 가계동향조사 표본설계에 그대로 반영된다는 것을 전제로 한 것은 아니지만 좀 더 구체적인 검토를 통해 향후 표본설계에 반영하는 경우 조사의 효율성 제고 차원에서 큰 도움이 될 것으로 판단되어 제안하는 것이다.

3) 추정식 및 분산추정식 산출방안 연구

현행 가구부문 경상조사를 위한 표본가중치는 지역 층 내에서 동일한 자체 가중치(equal & self weights)로 정의되고 있다. 정의된 표본 가중치는 calibration 과정을 통해 최종결과 분석을 위한 최종가중치로 보정된다. 경제활동인구조사의 경우

calibration을 위해 각 개인의 성별, 거주 시·도, 거주 동부/읍·면부, 연령 변수가 사용되었다. 이런 방식으로 계산된 최종가중치를 각 조사시점에 조사된 가구 및 가구원 데이터에 부여하여 조사 시점별 각종 추정량을 산출하고 있다.

산출된 추정량들의 분산 추정량으로는 반복분산추정 방안 중 하나인 잭나이프 분산추정 방안이 사용되고 있다. 이를 위하여 이상추출(two-phase sampling) 구조를 가지고 있는 현행 가구부문 경상조사 표본에 대한 가정을 하고, 주어진 가정 하에서 잭나이프 반복 가중치를 사용하여 분산 추정량을 계산하고 있다.

본 연구에서는 먼저 해외에서 사용되고 있는 연동 모형 하에서의 추정량들과 그 성질을 살펴보고 이를 통해 현행 통계청에서 사용하고 있는 사후층화추정량의 타당성을 평가하였다. 이를 위하여서 현행의 연동모형 하에서의 고려된 추정량들의 장·단점을 비교, 분석하였다. 이와 더불어 사용되고 있는 잭나이프 분산 추정량의 산출과정을 검토하여 그 문제점을 파악하고 해결 방안을 제시하였다. 또한 분산추정을 위해 고려된 가정이 만족되지 않을 때, 즉 이상표본 추출법을 통하여 추출된 표본의 가구 및 가구원 가중치가 동일하지 않을 때 적용할 수 있는 추정량과 이의 분산 추정을 위한 반복추정방안을 제시하였다.

마지막으로 지난 3년의 조사결과를 이용하여 패널조건화(panel conditioning)로 인한 편향, 즉 각 패널이 동일한 조사에 노출된 빈도에 따라서 발생하는 체계적인 편향의 유·무를 파악하였다. 이를 통하여 현행 사용되고 있는 추정량의 통계적 타당성 및 가능한 추정방법 개선 방안을 제시하였다.

II. 연동 모형 관련 기존 연구결과 검토

1. 기존 통계청 연동모형 연구 결과 요약

본장에서는 과거 통계청의 연동모형 관련 연구와 외국의 연동모형 사례를 분석하고 이를 기초로 추후 고려할 수 있는 연동모형을 제안하고자 한다. 통계청의 연동모형 관련 연구내용은 2002년 조사관리과에서 실시된 연구결과를 토대로 정리하였으며 외국 사례 분석을 위해서는 연동모형을 사용하고 있는 주요 선진국 자료들을 참조하였다.

통계청에서는 2002년 이전 표본관리 편의성을 고려해 한번 표본으로 선정된 가구를 다음 표본 개편까지 계속 조사하는 고정표본 관리방법을 사용하였으며, 이를 개선하기 위해 연동모형을 도입하기로 결정하고 관련 사전 연구를 수행하였다. 기존의 고정표본모형에서의 응답자는 5년간 각종 통계조사에 응하여야 하는 응답부담을 안게 되고, 표본은 시간이 지남에 따라 모집단의 변화를 반영하지 못해 노후화되는 문제를 가지고 있었다. 이를 해소하는 동시에 전면적 표본개편에 따른 신규계열간의 괴리현상을 최소화하기 위해서 매월 표본의 일부를 교체하여 주는 표본관리 기법의 도입을 고려하였고 이를 위해 1999년 11월부터 2000년 3월까지의 이론적 연구, 2000년 5월부터 2002년 6월까지 시험조사, 그리고 2004년 5월부터 인천과 경남지역에 대해 선행조사를 걸쳐 2005년부터는 전국에 걸쳐 본격적인 연동표본제가 실시됐다.

본 절에서는 2000년 5월부터 시행된 시험조사 내용을 바탕으로 가능한 연동모형 방안을 검토한 통계청의 연구결과를 정리하였다. 통계청에서 고려한 연동모형은 크게 4-8-4(변형 4-8-4)와 6-6-6방안이며 이러한 방안들의 평가를 위해서는 기본적으로

응답률, 가계부 회수율이 고려되었다. 특별히 응답회수 편향지수(Time in Sample Bias Index; TIS)를 사용하여 응답회수별 편향의 가능성 여부를 판단하였다.

연동방안의 적용을 통해 이루어진 두 조사는 경제활동인구조사와 도시가계조사로서 두 조사를 위해 고려된 시험조사의 기간, 조사규모 그리고 조사지역은 <표 2-1> 및 <표 2-2>와 같다. 또한 각 조사별 고려된 적용 연동모형은 경제활동인구조사의 경우, 4-8-4, 변형 4-8-4 그리고 6-6-6의 세 방안으로서 각 방안별 내용은 다음과 같다.

- 4-8-4 : 4개월 조사, 8개월 조사중지, 4개월 재조사 후 표본에서 제외
- 변형 4-8-4 : 4-8-4와 비슷하며 단지 1개의 조사구를 2개의 조사구로 간주
- 6-6-6 : 6개월 조사, 6개월 조사중지, 6개월 재조사 후 표본에서 제외

도시가계조사의 경우 18개월 조사 후 표본에서 제외하고 하나의 방안만을 고려하였다. 두 조사별 시험조사에서 사용된 연동방안에 대한 보다 자세한 내용은 다음과 같다.

<표 2-1> 경제활동 연동모형 시험조사 규모 및 지역 (단위: 가구)

기간	조사규모	조사지역
2000.05~2000.12	560	수원, 경기 대전, 충남, 대구, 경북(6개 지역)
2001.01~2001.12	840	+ 서울, 경남, 강원(9개 지역)
2002.01~2002.06	280	서울, 경남, 강원(3개 지역)

<표 2-2> 도시가계 연동모형 시험조사 규모 및 지역 (단위: 가구)

기간	조사규모	조사지역
2000.05~2000.12	180	서울, 대전, 전북(3개 지역)
2001.01~2001.12	480	+ 수원, 인천, 광주, 부산, 충북(8개 지역)
2002.01~2002.06	420	서울, 전북, 수원, 인천, 광주, 부산, 충북(7개 지역)

(1) 경제활동인구조사: 4-8-4, 변형 4-8-4, 6-6-6

- 4-8-4 : 4개월 조사, 8개월 조사중지, 4개월 재조사 후 표본에서 제외
- 변형 4-8-4 : 4-8-4와 비슷하며 단지 1개의 조사구를 2개의 조사구로 간주
- 6-6-6 : 6개월 조사, 6개월 조사중지, 6개월 재조사 후 표본에서 제외

(2) 도시가계조사

- 18개월 조사 후 표본에서 제외

1) 경제활동인구조사

경제활동인구조사를 위한 4-8-4, 변형 4-8-4, 6-6-6 연동모형의 효율을 비교하기 위해 <표 2-3>의 4지역에서 시험조사를 시행하였다. 아래 표본 조사구는 전년 동월 대비 50% 중복 조사구에 대하여 동일 가구와 다른 가구의 비중은 각각 89.2%, 10.8%로 나타내고 있다. 즉, 이 시험조사에 참여한 표본 가구 중 89.2%는 작년 경제활동인구조사에 참여한 가구이고 10.8%는 작년 경제활동인구조사에 참여하지 않은 가구이다. 따라서 이 시험조사에 사용된 표본 조사구는 지속표본과 진입표본이 모두 존재하는 것으로 파악된다.

<표 2-3> 경찰 인구조사 조사지역 및 표본구성

지역	조사기간	조사구수	동일가구	다른가구	전체
충남, 대전	2000.05~2001.12	96	864	83	947
서울	2001.01~2002.06	36	301	20	321
대구, 경북, 경기, 수원	2000.05~2001.12	128	1190	142	1332
경남, 강원	2001.01~2002.06	48	402	89	491
전체		308	2757	334	3091

(1) 연동모형별 응답률

<표 2-4> 4-8-4연동시스템의 응답횟수별 응답률

지역 \ 응답횟수	1	2	3	4	5	6	7	8
대구, 경북, 경기, 수원	97.6	97.7	97.8	97.7	100	100	100	100
경남, 강원	96.7	96.8	97.3	97.6	100	100	100	100
전체	97.3	97.4	97.7	97.7	100	100	100	100

<표 2-5> 6-6-6연동시스템의 응답횟수별 응답률

지역 \ 응답횟수	1	2	3	4	5	6	7	8	9	10	11	12
충남, 대전	100	100	100	100	100	100	100	100	100	100	100	100
서울	96.5	95.9	95.5	95.5	96.2	95.7	98.2	97.9	97.4	96.6	100	100
전체	99.0	99.0	99.0	98.7	98.9	98.7	99.5	99.5	99.4	99.2	100	100

<표 2-4>와 <표 2-5>는 4-8-4 연동모형과 6-6-6 모형하에서의 응답률 패턴을 각각 지역별로 나타내고 있다. 4-8-4 연동모형의 경우, 1~4회 응답횟수별 응답률은 97.3~97.7%를 나타내고 있으며, 8개월 비조사 기간 후인 5회째 응답횟수 이상인 경우 100%의 응답률을 나타내고 있다. 6-6-6 연동모형의 경우, 응답횟수가 1회일 때에도 97% 이상의 높은 응답률을 나타내고 있으며, 시간이 지날수록 높은 회수율을 나타내고 있다. 특히 일정기간 조사 이후 조사를 중지한 뒤 재조사의 경우 더 높은 응답률을 나타내고 있다.

2) 도시가계 조사

<표 2-6> 18개월 연속조사 연동모형 : 응답횟수별 가계부 회수율

응답횟수 지역	1	2	3	4	5	6	7	8	9	10
전체	42.0	55.3	59.3	60.8	62.1	62.7	63.9	64.4	63.6	64.2
서울	24.1	32.7	40.4	39.6	40.4	41.0	42.4	39.6	44.8	44.3
부산	42.7	50.0	55.3	59.0	59.8	63.9	65.1	67.2	67.3	71.1
인천	33.3	55.7	58.4	59.8	63.4	65.8	66.7	71.0	63.9	66.7
광주	49.5	62.0	61.5	64.3	65.9	64.4	67.6	68.1	64.9	63.3
대전	47.9	64.1	68.5	69.0	73.8	69.3	72.7	77.3	75.0	76.1
경기	33.7	48.1	53.8	58.7	60.8	63.0	65.0	65.5	60.8	65.0
충북	53.9	66.3	70.5	69.1	68.3	72.2	70.3	68.8	70.4	70.5
전북	51.8	66.4	69.2	71.1	70.8	70.2	71.1	72.0	73.1	72.1

<표 2-7> 경상 도시가계조사의 가계부 회수율

조사월 (응답 개월)	97.10 (1)	97.11 (2)	97.12 (3)	98.1 (4)	98.2 (5)	98.3 (6)	98.4 (7)	98.5 (8)	98.6 (9)	98.7 (10)
회수율(%)	54.3	67.8	71.1	75.0	76.6	78.1	79.1	80.2	79.5	80.0

<표 2-6>와 <표 2-7>은 도시가계조사의 가계부 회수율을 나타내고 있다. 응답 횟수별 가계부 회수율은 첫 1개월째 42.0%, 2개월째 55.3%, 3개월째 59.3%로 조사 시작 시점부터 3개월간은 증가하는 추세를 보이지만, 3개월이 지난 후 부터는 증가폭이 감소하고 있다. <표 2-7>과 비교하여 본 결과 과거에 시행한 경상 도시가계 조사에서도 시험조사와 마찬가지로 3개월 이후 안정화 되고 있으며 경상 도시가계 조사와 비교하여 시험조사에서의 회수율이 낮은 이유는 회수율이 낮은 지역이 상대적으

로 많이 추출되었기 때문으로 고려된다. 참고로 서울, 인천, 경기에서는 회수율이 낮고, 전북, 충북, 대전지역에서는 회수율이 높은 것으로 파악되고 있다.

3) 연동시스템 시험조사 연구 결과 요약 및 한계점

(1) 응답률 분석

<표 2-8>과 <표 2-9>는 시험조사를 통해 산출된 경제활동인구조사와 도시가계조사의 응답횟수별 응답률 및 가계부 회수율을 나타내고 있다. 표본교체에 따른 경제활동인구조사의 응답률은 차이가 거의 없는 것으로 나타나고 있고, 가계조사의 경우는 표본교체에 따른 회수율은 처음 3개월간은 회수율 차이가 현저하나 그 이후에는 소폭으로 증가하는 것으로 나타났고 이는 과거 진행되었던 가계조사에서의 회수율과 유사하다.

<표 2-8> 경제활동인구조사 응답률

응답횟수	1	2	3	4
응답률	97.22	97.42	97.51	97.51

<표 2-9> 도시가계조사 가계부 회수율

응답횟수	1	2	3	4	5	6	7	8	9
회수율	36.5	51.1	57.7	58.2	60.5	55.9	59.4	61.4	63.3

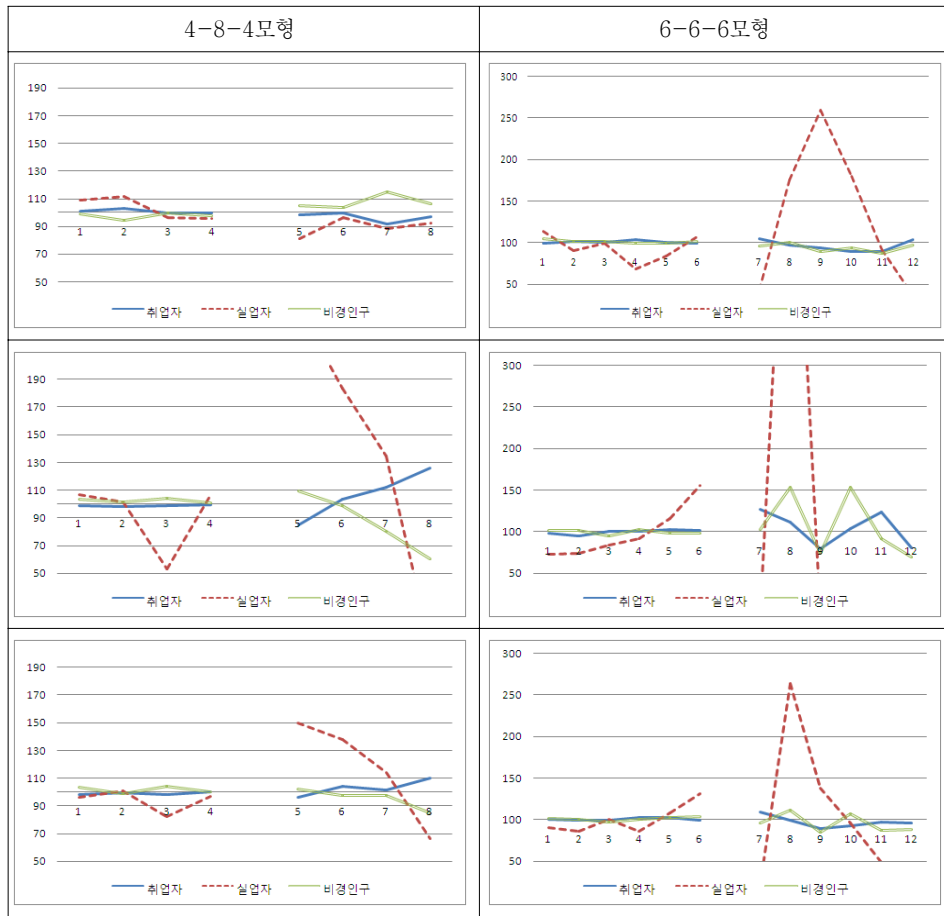
(2) 응답횟수별 편향지수 분석

응답횟수에 따른 단순한 응답률 분석이 아닌 자료의 질을 평가하기 위하여 3변수(취업자, 실업자, 비경제활동인구)를 이용하여 응답횟수별 편향지수 분석을 실시하였다. 이 때 정의되는 응답횟수별 편향지수는

$$I_{h,i} = \frac{E(X_{h,i})}{E(\sum_{j=1}^m X_{h,j}/m)} \times 100$$

로서 이는 해당 횟수 추정량의 다른 횟수 추정량들에 대한 상대적 크기를 나타낸다. [그림 2-1], <표 2-10> 그리고 <표 2-11>은 4-8-4 연동모형과 6-6-6 연동 모형 하에서 산출된 응답횟수별 편향지수를 나타내고 있다. 주어진 표를 통해 파악할 수 있는 특징들은 다음과 같다.

- 응답횟수 편향지수(TIS)는 지역별로 다른 경향을 나타내고 있어서 일률적으로 판단하기 어렵다.
- 응답횟수 편향지수는 실업률에서 특히 값이 많이 불안정하다.
- 4-8-4 모형의 경우 수원, 경기 대구, 경북의 취업자 편향지수는 크며 후반기에는 적은 경향을 보이지만, 경남 및 강원지역에서는 반대로 나타난다. 동일 연동모형이지만, 지역에 따라서 편향지수가 다른 경향을 보인다. 조사구 수가 적은 곳을 제외하고는 편향지수(TIS)는 91~104로 연동그룹에 따라 추정량이 크게 달라지지 않는다.
- 6-6-6 모형의 경우 대전, 충남지역의 실업자 편향지수는 첫 1회에서 과대, 그 이후 작게 나타나다가 6회째 다시 크게 나타나 일률적으로 증가하거나 감소하지는 않는 반면, 서울의 경우 응답횟수가 증가(1회→6회)할수록 과소추정에서 과대 추정되는 경향이 나타난다.



[그림 2-1] 경제활동인구조사 조사지역 및 표본구성

〈표 2-10〉 4-8-4모형에서 응답횟수별 편향지수(TIS)

응답 횟수	0005-0112(수원, 경기, 대구, 경북)				0101-0206(경남, 강원)				전체		
	적용월 (가구 수)	취업자	실업자	비경 인구	적용월 (가구 수)	취업자	실업자	비경 인구	취업자	실업자	비경 인구
1	17 (920)	101.178	109.222	99.072	15 (840)	98.816	106.850	103.660	98.284	96.320	103.234
2		103.095	111.293	94.676		98.267	101.315	101.456	99.358	100.940	98.838
3		99.603	96.533	99.645		98.478	53.313	104.340	98.351	82.159	103.856
4		99.963	95.506	97.639		99.148	106.547	100.726	99.804	96.721	100.277
5	5 (200)	98.454	81.064	105.272	3 (120)	85.056	241.923	109.183	96.211	149.839	102.397
6		99.913	96.020	103.874		103.368	183.333	98.899	103.892	137.998	97.570
7		91.640	88.184	114.722		112.004	134.622	80.405	101.762	114.148	97.216
8		96.936	92.047	106.622		126.024	0.000	60.602	110.221	66.610	84.980

〈표 2-11〉 6-6-6모형에서 응답횟수별 편향지수(TIS)

응답 횟수	0050-0112(대전, 충남)				0101-0206(서울)				전체		
	적용월 (가구 수)	취업자	실업자	비경 인구	적용월 (가구 수)	취업자	실업자	비경 인구	취업자	실업자	비경 인구
1	15 (1020)	99.791	113.893	104.433	13 (900)	98.241	73.131	101.324	100.232	90.787	101.210
2		100.973	90.378	101.176		95.006	74.371	101.250	98.720	85.871	100.593
3		100.344	99.263	101.534		100.842	83.538	94.932	99.640	100.331	97.261
4		103.149	67.903	99.059		100.691	91.253	102.956	103.038	86.026	100.386
5		100.869	83.986	99.245		102.070	115.703	98.545	102.242	107.811	102.113
6		99.288	107.442	101.837		101.185	155.410	97.730	99.516	131.530	103.253
7	3 (180)	104.511	42.569	95.891	1 (60)	127.388	0.000	102.053	108.748	22.111	95.631
8		96.694	175.486	100.010		111.465	685.714	153.080	99.040	263.984	111.937
9		94.222	259.175	89.798		78.821	0.000	73.478	89.085	137.407	84.360
10		89.823	180.019	94.050		103.503	0.000	153.080	92.140	96.194	107.483
11		89.456	91.659	86.650		123.769	0.000	91.848	97.021	48.834	87.142
12		103.221	36.764	97.182		80.613	0.000	68.886	96.182	19.096	88.154

연동표본 비교를 위해 시행된 시험조사는 그 결과를 비교하기에 많은 한계점을 가지고 있다. 기본적인 응답률 분석 이외의 결과, 특별히 추정량에 대한 분석의 경우, 시험조사에 사용된 표본크기(특별히 지역별, 응답횟수별 표본크기)가 너무 적다. 또한 4개월이라는 짧은 시간을 통한 종단면적 분석 자체가 어렵다는 한계도 있다. 다만 응답률분석을 통해 파악된 내용을 볼 때 고려된 두 모형이 크게 차이가 나지 않을 것으로 파악된다.

4) 현행 1/36 연동방안

현재 통계청에서는 1/36 연동방안을 경제활동인구조사를 위해 사용하고 있다. 이전 연구에서 실제 이 연동방안을 타 연동방안과 비교하지는 않았다. 그러나 시험조사결과 고려된 연동방안별(4-8-4, 6-6-6) 응답률 차이가 발생하지 않는다는 것, 응답자의 부담을 경감시키며 시계열적 안정성을 유지, 또한 조사 업무수행의 용이함과 같은 현실적인 문제를 고려하여 1/36 연동방안을 채택한 것으로 보인다. 개략적으로 1/36 연동모형 관련 내용을 정리하면 다음과 같다.

(1) 목적

- 연동표본제 정착으로 응답자 부담경감 및 안정적 시계열 자료 생산
 - 기존 고정표본에서는 한 응답자는 5년간 각종 통계조사에 응하여야 하는 응답 부담을 가지게 됨. 또한, 5년에 한 번씩 표본을 추출하기 때문에 시간이 지남에 따라서 모집단을 반영하지 못함. 따라서 5년 후 새로운 표본을 추출하게 되면 시계열 자료가 불안정해짐.

(2) 교체방법

- 표본으로 선정된 가구가 36개월 동안 조사에 참여하고, 이후 표본에서 완전히 제외

(3) 표본특징

- 전월 대비 중복률 35/36, 전년 대비 중복률 24/36
 - 전월대비 전년대비 중복률이 높음.
 - 두 가지 표본의 단점을 보완시키는 방법이 연동표본인데 다른 나라의 4-8-4, 1/8, 1/6보다 조사기간도 길고, 연동그룹도 많음.

(4) 연동그룹의 구조

연동표본은 모든 조사구를 9개의 그룹으로 나누어 구성되고, 1개의 그룹은 181개의 조사구로 구성됨. 하나의 조사구는 4개의 구역으로 이뤄져 있는데, 1개의 구역은 5가구로 이루어져 있음. 따라서 1개 그룹의 조사구는 4개월에 걸쳐 인접 구역으로 연동교체 되므로 모든 조사구, 가구의 연동교체에는 36개월이 소요됨.

2. 외국 연동 교체 방안 비교 연구

본 절에서는 해외에서 사용되고 있는 여러 가지 연동방안들을 소개하고 그 특징들을 살펴보고자 한다. 고려되는 연동방안들은 크게 7가지 경우로 구분할 수 있다.

1) 1/8교체 방안

- 교체방법: 한 응답자가 8번 연속적으로 표본에 포함된 후 표본에서 제외
- 표본특징: 교체주기에 따라서 표본 특징 달라짐

(1) 1/8교체 방안 ①

- 교체주기: 매달
- 교체방법: 표본에 포함된 가구는 8개월 동안 조사에 참여한 후 교체
- 전월대비 중복표본 비율: 7/8, 전년대비 중복표본 비율: 0
- 사용국가: 오스트레일리아, 스웨덴

(2) 1/8교체 방안 ②

- 교체주기: 분기
- 교체방법: 표본에 포함된 가구는 2년 동안 8번 조사에 참여한 후 교체
- 이전 분기대비 중복표본 비율: 7/8, 전년대비 중복표본 비율: 4/8
- 사용국가: 오스트리아

2) 1/6교체 방안

- 교체방법: 한 응답자가 6번 연속적으로 표본에 포함된 후 표본에서 제외

- 표본특징: 교체주기에 따라서 표본의 특징이 달라짐

(1) 1/6교체 방안 ①

- 교체주기: 매달
- 교체방법: 표본에 포함된 가구는 6개월 동안 조사에 참여한 후 교체
- 전월대비 중복표본 비율: 5/6, 전년대비 중복표본 비율: 0
- 사용국가: 캐나다

(2) 1/6교체 방안 ②

- 교체주기: 분기
- 교체방법: 표본에 포함된 가구는 18개월 동안 6번 조사에 참여한 후 교체
- 이전 분기대비 중복표본 비율 : 5/6, 전년대비 중복표본 비율: 2/6
- 사용국가: 스페인

(3) 1/6교체 방안 ③

- 교체주기: 반기
- 교체방법: 표본에 포함된 가구는 3년 동안 6번 조사에 참여한 후 교체
- 이전 반기대비 중복표본 비율: 5/6, 전년대비 중복표본 비율 : 4/6
- 사용국가: 베네수엘라

3) 1/5교체 방안

- 교체방법: 한 응답자가 5번 연속적으로 표본에 포함된 후 표본에서 제외
- 교체주기: 분기
- 교체방법: 표본에 포함된 가구는 15개월 동안 5번 조사에 참여한 후 교체
- 이전 분기대비 중복비율: 4/5, 전년대비 중복비율: 1/5
- 사용국가: 멕시코

4) 1/4 교체 방안

- 교체방법: 한 응답자가 4번 연속적으로 표본에 포함된 후 표본에서 제외
- 표본특징: 교체주기에 따라서 표본의 특징이 달라짐

(1) 1/4 교체 방안 ①

- 교체주기: 분기
- 교체방법: 표본에 포함된 가구는 1년 동안 4번 조사에 참여한 후 교체
- 이전 분기대비 중복비율: 3/4, 전년대비 중복비율: 0
- 사용국가 : 포르투갈, 필리핀

(2) 1/4 교체 방안 ②

- 교체주기: 매년
- 교체방법: 표본에 포함된 가구는 4년 동안 4번 조사에 참여한 후 교체
- 전년대비 중복비율: 3/4, 연속되는 3년 동안 중복비율: 2/4
- 사용국가: 독일, 그리스

5) 1/2 교체 방안

- 교체방법: 한 응답자가 2번 연속적으로 표본에 포함된 후 표본에서 제외
- 교체주기: 반기
- 교체방법: 표본에 포함된 가구는 1년 동안 2번 조사에 참여한 후 교체
- 이전반기대비 중복비율: 1/2, 전년대비 중복비율: 0
- 사용국가: 터키

6) 2-2-2 교체 방안

- 교체방법: 한 응답자가 처음 2회 동안 연속적으로 표본에 포함된 후 다음 2회동안 표본에서 제외되었다가 그 다음 2회 동안 다시 표본에 속한 후 완전히 표본에서 제외되는 표본조사 방식
- 교체주기: 분기
- 이전 분기대비 중복비율: 2/4, 전년대비 중복비율 : 0
- 사용국가: 노르웨이

7) 4-8-4 교체 방안

- 교체방법 : 한 응답자가 처음 4회 동안 연속적으로 표본에 포함된 후 다음 8회 동안 표본에서 제외되었다가 그 다음 4회 동안 다시 표본에 속한 후 완전히 표본에서 제외되는 표본조사 방식
- 교체주기: 매월
- 전월대비 중복비율: 6/8, 전년대비 중복비율: 4/8
- 사용국가: 미국, 일본

각 국가에서 현재 사용되고 있는 연동모형을 선택한 이유는 매우 다양하다. 대부분의 경우, 응답자별 조사자 부담, 비용과 같은 현실적인 문제들과 연결되어 있다. 연동모형에서 중복률이 높을수록 시간에 따라 산출되는 추정량이 안정적이거나 각 조사대상자의 조사 참여 기간이 길어짐으로 패널 조건화와 같은 비표본오차가 발생할 가능성은 높아진다. 주요 선진국에서는 일반적으로 현실적인 문제들을 고려한 연동방안을 채택하고, 주어진 연동방안 하에서 효율성을 높일 수 있는 추정량을 연구 개발하여 사용하는 전략을 취하고 있는 것으로 파악된다.

III. 연동표본추출을 위한 표본설계

1. 기존 표본설계의 타당성 검증

본 장에서는 현행 경제활동인구조사 표본설계의 타당성 검증을 위해 우선 무응답에 따른 편향 발생 가능성을 점검해 보기 위해 경제활동인구조사와 가계동향조사의 응답률을 다양한 관점에서 심층적으로 분석하였다. 아울러 현행 표본설계의 효율성을 살펴보기 위해 주요 변수별 집락효과 분석을 통해 조사구당 20가구를 조사하는 현재의 집락추출법의 적절성을 평가한다. 한편 현행 경제활동인구조사 표본설계에 있어 핵심적인 역할을 하는 층 내 정렬변수의 적절성을 검토하기 위해 2005년 인구주택총조사 당시의 조사구별 정보와 현행 표본 조사구에서 조사되는 변수들 간의 연관성을 분석하였다.

1) 응답 패턴 분석

경제활동인구조사는 현재 표본 조사구에서 4개 구역의 20가구를 표본으로 추출하고, 조사 대상 가구에서 만 15세 이상 가구원을 대상으로 조사를 실시하며, 가계동향조사는 표본 조사구에서 2개 구역의 10가구를 표본으로 선정하고, 가구 기준으로 자료를 수집하고 있다. 현행 경제활동인구조사 및 가계동향조사에서 무응답에 따른 편향 발생 가능성을 검토하기 위해 여기서는 가구 수준 뿐 아니라 가구원 수준의 무응답 현황에 대해 분석해 보고자 한다. 실제 무응답 패턴에 있어서 연도별로 큰 차이가 없으므로, 여기서는 가장 최근 자료인 2010년 자료 위주로 분석하였다.

(1) 경제활동인구조사 응답 현황

경제활동인구조사는 연동표본 방식이 적용되고 있기 때문에 연동에 따른 표본 가구 변동이 발생하는 시점에는 무응답 가구와 연동에 따른 가구 변동이 혼재되어 분석이 용이하지 않을 수 있다. 따라서 여기서는 순수한 무응답에 의한 가구 응답 패턴을 분석하기 위해 연동으로 인한 조사구 변동이 없는 4, 8, 12월 조사대상 응답 가구 및 가구원을 분석대상으로 하였다. <표 3-1>을 보면 2010년 경제활동인구조사 대상인 32,580가구 중에서 3개월 동안 평균 30,102가구(92.4%)가 응답하였으며, 가구원 기준으로는 평균 67,821명(가구당 약 2.25명)이 응답한 것으로 나타났다.

<표 3-1> <경찰> 층별 응답 현황

층	표본	가구			가구원		
		4월	8월	12월	4월	8월	12월
전국	32,580	30,247	30,004	30,055	67,977	67,766	67,719
서울	4,140	3,807	3,730	3,773	9,053	8,911	8,962
부산	2,160	2,052	2,000	1,997	4,892	4,755	4,679
대구	1,800	1,678	1,635	1,646	3,878	3,812	3,830
인천	1,980	1,771	1,747	1,750	4,180	4,165	4,121
광주	1,620	1,359	1,404	1,435	3,111	3,231	3,339
대전	1,620	1,516	1,488	1,483	3,487	3,456	3,427
울산	1,440	1,370	1,357	1,365	3,140	3,107	3,077
경기/동부	2,700	2,473	2,447	2,435	5,988	5,936	5,916
경기/읍면	1,260	1,151	1,158	1,170	2,571	2,635	2,662
강원/동부	900	825	821	818	1,757	1,751	1,778
강원/읍면	900	872	881	872	1,828	1,833	1,819
충북/동부	900	829	825	812	1,821	1,823	1,811
충북/읍면	900	869	872	874	1,819	1,842	1,820
충남/동부	900	811	807	828	1,820	1,825	1,890
충남/읍면	900	848	847	838	1,812	1,808	1,811
전북/동부	900	838	809	805	1,950	1,888	1,873
전북/읍면	720	708	708	711	1,401	1,401	1,398
전남/동부	900	850	850	848	1,790	1,828	1,822
전남/읍면	900	851	865	856	1,641	1,678	1,677
경북/동부	1,080	1,018	1,025	1,017	2,168	2,233	2,193
경북/읍면	1,080	1,053	1,050	1,046	2,091	2,061	2,050
경남/동부	1,080	959	948	966	2,127	2,140	2,155
경남/읍면	900	855	843	836	1,731	1,717	1,707
제주/동부	540	511	514	505	1,165	1,167	1,149
제주/읍면	360	373	373	369	756	766	753

경제활동인구조사와 같은 연속조사의 경우 동일 조사단위가 매월 지속적으로 조사에 참여하는지 여부가 조사의 품질을 결정하는 중요한 요인이 될 수 있다. 따라서 연도별 가구 및 가구원 월별 응답률을 살펴보기 위해, 우선 경제활동인구조사의 조사대상 가구 중 매년 연동에 의해 교체되는 조사구에 포함되는 경우를 제외하고, 연도별로 연간 12개월 동안 매월 지속적으로 조사가 되어야 할 표본 가구 규모를 파악한 결과 2008~2010년 동안 매년 2만3천~2만5천 가구인 것으로 나타났다. 여기서는 이들 연동에 따른 교체가 발생하지 않는 조사구의 가구를 기준으로 가구 응답패턴을 분석하기로 한다.

또한 조사구에 변동이 없더라도 표본 조사구에 가구 전입 및 전출이 발생하기 때문에 12개월 동안의 가구응답률을 정확히 산출하기 위해, 1~11월에 전출된 가구와 2~12월에 표본에 추가된 가구를 제외하고, 실제 1년 동안 매월 조사가 되어야 할 표본 가구를 대상으로 응답률을 산출한 결과는 <표 3-2>와 같다. 여기서 가구 기준 응답률 계산에 있어서 분모가 되는 1월 응답 가구는 2008년에는 17,918가구, 2009년과 2010년에는 각각 18,031가구, 17,802가구 인 것으로 나타났고, 이 중 12개월 모두 응답한 가구는 2008년 91.5%, 2009년과 2010년은 각각 97.1%인 것으로 나타나 철저한 표본관리를 통해 높은 응답률을 유지하고 있는 것으로 보인다.

가구원 응답률은 1월에 응답한 가구내 가구원을 기준으로 산출하였으며, 대상 가구 내에 가구원 전출/입을 제외하면, 기준 가구(1월 응답)의 가구원 중에서 12개월 모두 응답한 가구원 응답률은 2008년에는 79.0%, 2009년과 2010년에는 각각 84.2%, 83.8%인 것으로 나타났다.

<표 3-2> <경향> 응답 가구 및 가구원 현황

연도 (제외한 연동그룹)	가구(가구)				가구원(명)		
	연간	1월 기준	12회 응답	응답률 (%)	연간	12회 응답	응답률 (%)
2010년 (7~9)	23,266	17,802	17,281	97.1	44,592	37,365	83.8
2009년 (4~6)	23,741	18,031	17,508	97.1	45,090	37,977	84.2
2008년 (1~3)	25,192	17,918	16,394	91.5	44,325	35,038	79.0

그러나 동일 가구 전출/입 외에 다양한 이유로 표본에 이탈/진입하는 가구가 있기 때문에 정확한 가구원 기준 응답률 분석에는 한계가 있다는 점에 유의할 필요가 있다. <표 3-3>을 보면, 2010년의 경우 1~11월에 마지막으로 응답한 가구가 3,236가

구이지만 이들 중 전출 가구로 파악되는 가구는 2,762가구이고 불응과 부적적으로 제외된 가구 이외에도 84가구는 응답하지 않은 사유를 정확히 파악할 수 없는 경우에 해당한다. 한편 2~12월에 처음으로 응답하기 시작한 가구는 2,814가구인데 이 중 전입 가구로 파악된 가구는 2,584가구이고 나머지 230가구는 기타 사유로 표본에 새로 유입된 것으로 파악된다. 특히 2008년의 경우 표본에서 유출된 가구 중 무응답 사유를 파악할 수 없는 경우에 해당하는 가구가 945가구이고, 유입된 가구 중 전입 이외 사유에 해당하는 가구가 1,148가구인 것으로 나타났다. 사유를 알 수 없어 기타로 구분된 경우는 구역 조정으로 인한 가구 코드 변경이나 코드 입력 오류 등에 의해 것으로 추측되며, 2008년에 비해 2009년이나 2010년에는 이런 사례가 현격하게 감소했다는 것은 최근 보다 체계적인 표본 관리가 이루어지고 있다는 것을 보여주는 긍정적인 현상으로 보인다.

〈표 3-3〉 <경찰> 가구 이탈/전입 현황

연도 (제외한 연동그룹)	1-11월 마지막 응답						2-12월					
		전출	전출외	불응	제외	기타	처음 응답	전입	전입외	조사구 변동	가구 추가	기타
2010년 (7~9)	3,236	2,762	474	379	11	84	2,814	2,584	230	7	0	223
2009년 (4~6)	3,306	2,884	422	350	19	53	3,034	2,835	199	177	1	21
2008년 (1~3)	4,391	3,283	1,108	135	28	945	4,336	3,188	1,148	300	61	787

아울러 경제활동인구조사와 같은 연속조사에 있어서는 안정적인 자료 생성을 위해서는 조사대상 가구가 얼마나 지속적으로 조사에 응답하고 있는지를 파악하고, 응답률을 높일 수 있는 방안을 모색하는 것이 중요할 것이다. 따라서 2010년 표본 가구 중 연동 교체 대상과 전출/입 가구를 제외하고, <표 3-2>에 제시된 2010년 1월에 응답한 17,802가구(이론상 이들 가구는 12개월 동안 계속 조사가 되어야 함)를 기준으로 2010년의 월별 응답률을 분석한 결과는 <표 3-4>와 같다. 일반적인 패널조사에서 발생하는 패턴과 유사하게 1월부터 시간이 흐름에 따라 매달 응답률은 점점 감소하는 경향이 나타나고 있다. 2010년 1월 표본가구를 100%로 하였을 때, 12월에는 그 중 98.0%가 응답하는 것으로 나타나, 1년 동안 발생하는 표본 마모율은 2% 수준으로 상당히 낮은 것으로 보인다. 앞에서 설명한 대로 <표 3-4>의 결과는 가구 기준으로 무응답이외에 전출입이나 부적격 제외 등에 해당하는 경우는 무응답률 계산에 반영하지 않은 결과라는 점에 유의하기 바란다.

<표 3-4> <경향> 2010년(1월 기준) 총 및 월별 응답 가구수 (단위: 가구)

총	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	17,802	17,741	17,661	17,618	17,583	17,540	17,519	17,484	17,445	17,452	17,450	17,440
	100.0	99.7	99.2	99.0	98.8	98.5	98.4	98.2	98.0	98.0	98.0	98.0
서울	2,158	2,140	2,116	2,105	2,094	2,083	2,081	2,083	2,072	2,078	2,079	2,082
부산	1,201	1,198	1,196	1,193	1,187	1,185	1,183	1,177	1,169	1,169	1,171	1,170
대구	972	971	967	967	967	967	962	959	957	959	959	958
인천	1,037	1,035	1,031	1,024	1,023	1,022	1,016	1,010	1,008	1,007	1,008	1,007
광주	819	811	798	795	798	794	794	790	789	790	791	790
대전	870	868	865	860	858	852	853	855	853	852	850	849
울산	816	815	811	811	808	806	806	803	804	806	804	804
경기/동부	1,386	1,378	1,373	1,372	1,371	1,368	1,367	1,362	1,357	1,355	1,355	1,355
경기/읍면	674	669	668	667	667	662	663	662	661	661	662	661
강원/동부	488	487	486	487	486	486	486	486	486	486	486	486
강원/읍면	536	536	536	535	535	535	535	535	535	535	535	535
충북/동부	471	470	467	465	465	463	459	457	457	458	458	457
충북/읍면	548	548	547	546	545	545	545	545	545	545	545	544
충남/동부	481	478	477	478	477	475	474	473	471	473	471	471
충남/읍면	522	521	518	518	516	516	514	514	510	510	510	510
전북/동부	499	496	496	492	490	488	486	488	487	487	487	487
전북/읍면	435	434	432	432	432	431	431	430	430	430	430	430
전남/동부	498	498	496	494	492	492	492	490	491	490	491	489
전남/읍면	526	525	523	523	524	523	523	522	522	523	523	522
경북/동부	608	607	607	607	606	605	604	604	604	603	602	603
경북/읍면	645	645	644	644	644	644	644	644	644	644	644	644
경남/동부	566	565	562	559	556	555	558	557	555	553	549	547
경남/읍면	512	512	511	511	511	510	510	509	509	509	510	509
제주/동부	298	298	298	298	296	298	298	295	295	295	296	296
제주/읍면	236	236	236	235	235	235	235	234	234	234	234	234

한편 2010년 1월에 응답한 가구 중 원칙적으로 12개월 동안 매월 조사에 응답해야 하는 가구를 대상으로 조사에 참여한 개월 수에 따라 분석해 본 결과는 <표 3-5>와 같다. 2010년의 경우 1월에 응답한 17,802가구 중(전출/입 제외)를 기준으로 총별로 1년 동안 응답한 횟수를 분석한 결과를 보면 전체적으로 2010년에 12회 모두 응답한 가구는 17,802가구 중 17,281가구로 97.1%가 2010년 1년 동안 매월 한 번도 무응답 없이 조사에 참여할 것으로 나타났다. 총별로 보면 12개월 동안 한 번도 누락되지 않고 응답한 가구비율이 서울과 광주가 94.4%로 다른 지역에 비해 조사 참여율이 낮은 것을 나타냈다.

참고로 <표 3-5>의 결과를 보면 지역별로 응답률이 대체적으로 97% 수준을 유지

하는 것으로 보아 실사과정에서 표본 가구 관리를 상당히 효율적으로 하고 있는 것으로 보인다. 향후 현재의 응답패턴을 계속 유지하거나 이를 보다 향상시키기 위해서는 이와 같은 조사 참여율을 계량적으로 나타낼 수 있는 지표를 산출하여 지속적으로 모니터링하는 것이 도움이 될 것으로 보인다.

〈표 3-5〉 〈경찰〉 2010년 총 및 응답횟수별 가구수

	합계	12회	11회	10회	9회	8회	7회	6회	5회	4회	3회	2회	1회
전체 (가구)	17,802	17,281 97.1	75 0.4	62 0.3	36 0.2	53 0.3	48 0.3	35 0.2	43 0.2	35 0.2	40 0.2	55 0.3	39 0.2
서울 가구 (%)	2,158	2,038 94.4	15 0.7	14 0.7	5 0.2	14 0.7	7 0.3	5 0.2	12 0.6	9 0.4	10 0.5	19 0.9	10 0.5
부산 가구 (%)	1,201	1,160 96.6	7 0.6	3 0.3	4 0.3	6 0.5	4 0.3	2 0.2	2 0.2	7 0.6	1 0.1	2 0.2	3 0.3
대구 가구 (%)	972	950 97.7	6 0.6	2 0.2	0 0.0	3 0.3	4 0.4	3 0.3	0 0.0	0 0.0	1 0.1	3 0.3	0 0.0
인천 가구 (%)	1,037	993 95.8	5 0.5	5 0.5	6 0.6	5 0.5	7 0.7	6 0.6	2 0.2	1 0.1	2 0.2	3 0.3	2 0.2
광주 가구 (%)	819	773 94.4	4 0.5	4 0.5	4 0.5	5 0.6	3 0.4	5 0.6	9 1.1	2 0.2	3 0.4	3 0.4	4 0.5
대전 가구 (%)	870	843 96.9	4 0.5	3 0.3	2 0.2	4 0.5	0 0.0	0 0.0	4 0.5	1 0.1	5 0.6	2 0.2	2 0.2
울산 가구 (%)	816	796 97.6	1 0.1	4 0.5	2 0.3	4 0.5	3 0.4	0 0.0	1 0.1	2 0.3	1 0.1	2 0.3	0 0.0
경기/동부 가구 (%)	1,386	1,346 97.1	6 0.4	2 0.1	4 0.3	4 0.3	6 0.4	3 0.2	2 0.1	0 0.0	3 0.2	4 0.3	6 0.4
경기/읍면 가구 (%)	674	655 97.2	5 0.7	2 0.3	0 0.0	0 0.0	2 0.3	1 0.2	2 0.3	1 0.2	0 0.0	3 0.3	4 0.6
강원/동부 가구 (%)	488	484 99.2	0 0.0	2 0.4	0 0.0	0 0.0	0 0.0	1 0.2	0 0.0	0 0.0	0 0.0	1 0.2	0 0.0
강원/읍면 가구 (%)	536	535 99.8	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.2	1 0.2	0 0.0	0 0.0
충북/동부 가구 (%)	471	454 96.4	1 0.2	3 0.6	0 0.0	0 0.0	2 0.4	4 0.9	1 0.2	1 0.2	2 0.4	2 0.4	1 0.2
충북/읍면 가구 (%)	548	544 99.3	1 0.2	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	1 0.2	1 0.2	1 0.2	0 0.0
충남/동부 가구 (%)	481	466 96.9	4 0.8	3 0.6	0 0.0	1 0.2	0 0.0	1 0.2	2 0.4	1 0.2	0 0.0	4 0.4	1 0.2
충남/읍면 가구 (%)	522	508 97.3	1 0.2	0 0.0	2 0.4	2 0.4	2 0.4	1 0.2	1 0.2	2 0.4	0 0.0	2 0.4	1 0.2
전북/동부 가구 (%)	499	484 97.0	1 0.2	0 0.0	2 0.4	2 0.4	0 0.0	0 0.0	2 0.4	2 0.4	3 0.6	0 0.0	3 0.6
전북/읍면 가구 (%)	435	429 98.6	0 0.0	0 0.0	1 0.2	0 0.0	1 0.2	0 0.0	1 0.2	1 0.2	0 0.0	2 0.5	0 0.0
전남/동부 가구 (%)	498	484 97.2	5 1.0	1 0.2	1 0.2	0 0.0	2 0.4	1 0.2	0 0.0	0 0.0	3 0.6	1 0.2	0 0.0
전남/읍면 가구 (%)	526	519 98.7	2 0.4	1 0.2	0 0.0	1 0.2	1 0.2	0 0.0	0 0.0	0 0.0	1 0.2	0 0.0	1 0.2
경북/동부 가구 (%)	608	601 98.9	2 0.3	0 0.0	0 0.0	1 0.2	1 0.2	1 0.2	0 0.0	1 0.2	0 0.0	0 0.0	1 0.2
경북/읍면 가구 (%)	645	644 99.8	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	1 0.2	0 0.0
경남/동부 가구 (%)	566	540 95.4	2 0.4	13 2.3	1 0.2	1 0.2	0 0.0	1 0.2	1 0.2	3 0.5	2 0.4	2 0.4	0 0.0
경남/읍면 가구 (%)	512	508 99.2	1 0.2	0 0.0	1 0.2	0 0.0	0 0.0	0 0.0	1 0.2	0 0.0	0 0.0	1 0.2	0 0.0
제주/동부 가구 (%)	298	293 98.3	2 0.7	0 0.0	1 0.3	0 0.0	2 0.7	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
제주/읍면 가구 (%)	236	234 99.2	0 0.0	0 0.0	0 0.0	0 0.0	1 0.4	0 0.0	0 0.0	0 0.0	1 0.4	0 0.0	0 0.0

한편, 앞에서 분석한 내용을 가구원 기준으로 산출해 보기 위해 2010년 1월에 응답한 17,802가구의 가구원을 대상으로 월별 응답여부와 연중 응답횟수를 분석해 보기로 한다. <표 3-6>을 보면 2010년 연간 응답 가구원 수는 월별 응답 가구수와 마찬가지로, 가구원수도 점점 감소하는 경향을 보이고 있다. 전체로 보면 1월 표본 가구 17,802가구에 포함된 41,410명의 가구원이 응답하였으나 이 중 12월에는 40,091명이 응답하였고, 평균적으로 매월 40,461명이 응답한 것으로 나타났다. <표 3-6>에 정리된 월별 응답 가구원수는 2010년 12개월 동안 전/출입한 가구원은 제외되었으나 그 외의 가구원 변동은 반영하지 않고 계산한 응답 가구원수라는 점에 유의하기 바란다.

표본 가구 내의 가구원 변동에 있어서 연령대별로 많은 차이가 있을 수 있다. 예를 들어 청년층의 경우 다양한 원인으로 표본 가구원 유입/유출이 자주 발생해 표본조사 대상 가구원 변동 가능성이 높기 때문에 연령대에 따라 응답패턴에 차이가 있을 수 있다. 이런 관점에서 2010년 전체 응답패턴 분석대상 41,410명을 5세 간격 연령대별로 구분하여 월별 응답자 수를 정리한 결과는 <표 3-7>과 같다. 분석결과 특이한 사항은 청년층에 해당하는 15~19세 및 20~24세 연령대의 경우 매월 지속적으로 감소하는 추세를 보이지만, 6-8월과 11-12월에 증가하는 현상을 보이고 있다. 이들 시점이 학생들의 경우 방학에 해당한다는 것이 이런 변동의 원인이 될 수 있을 것으로 추측은 되지만, 청년층의 이런 월별 변동이 청년층 경제활동인구 관련 통계 작성에 영향을 줄 수 있는 현상이기 때문에 이에 대해서는 추후 보다 심층적인 분석이 필요해 보인다.

〈표 3-6〉 〈경찰〉 2010년 총 및 월별 응답 가구원수 (단위: 명)

총	가구수	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	17,802	41,410	41,297	40,672	40,528	40,353	40,357	40,419	40,347	40,027	40,027	40,009	40,091
서울	2,158	5,266	5,226	5,160	5,130	5,085	5,059	5,068	5,083	5,034	5,040	5,039	5,045
부산	1,201	2,922	2,921	2,915	2,911	2,888	2,886	2,866	2,865	2,838	2,836	2,828	2,833
대구	972	2,351	2,354	2,335	2,324	2,323	2,330	2,313	2,295	2,272	2,278	2,273	2,274
인천	1,037	2,528	2,516	2,479	2,463	2,450	2,468	2,448	2,428	2,404	2,402	2,404	2,417
광주	819	1,953	1,933	1,885	1,879	1,880	1,872	1,877	1,864	1,874	1,878	1,884	1,875
대전	870	2,092	2,085	2,057	2,038	2,037	2,024	2,024	2,028	2,015	2,013	2,013	2,019
울산	816	1,908	1,906	1,874	1,869	1,858	1,864	1,867	1,861	1,860	1,867	1,857	1,873
경기/동부	1,386	3,526	3,513	3,450	3,452	3,446	3,452	3,471	3,456	3,413	3,405	3,419	3,444
경기/읍면	674	1,542	1,532	1,510	1,501	1,505	1,494	1,510	1,514	1,505	1,507	1,505	1,506
강원/동부	488	1,086	1,088	1,073	1,076	1,067	1,055	1,058	1,060	1,052	1,048	1,050	1,053
강원/읍면	536	1,185	1,187	1,167	1,167	1,160	1,160	1,164	1,158	1,145	1,139	1,132	1,136
충북/동부	471	1,065	1,069	1,051	1,045	1,038	1,039	1,044	1,043	1,039	1,034	1,032	1,028
충북/읍면	548	1,190	1,188	1,170	1,163	1,162	1,166	1,165	1,171	1,161	1,158	1,156	1,150
충남/동부	481	1,158	1,157	1,133	1,130	1,124	1,127	1,130	1,128	1,112	1,120	1,123	1,129
충남/읍면	522	1,186	1,189	1,163	1,156	1,149	1,148	1,147	1,147	1,135	1,134	1,135	1,139
전북/동부	499	1,192	1,182	1,165	1,163	1,159	1,158	1,155	1,155	1,139	1,142	1,142	1,147
전북/읍면	435	875	869	854	854	860	856	863	862	853	854	858	856
전남/동부	498	1,115	1,108	1,076	1,070	1,057	1,074	1,084	1,076	1,071	1,070	1,070	1,083
전남/읍면	526	1,030	1,025	1,011	1,012	1,011	1,008	1,008	1,010	1,001	1,004	1,001	1,000
경북/동부	608	1,393	1,382	1,354	1,345	1,343	1,344	1,356	1,356	1,336	1,335	1,331	1,332
경북/읍면	645	1,319	1,317	1,298	1,301	1,296	1,291	1,303	1,298	1,289	1,286	1,282	1,276
경남/동부	566	1,310	1,317	1,286	1,275	1,263	1,273	1,283	1,283	1,273	1,272	1,264	1,262
경남/읍면	512	1,052	1,054	1,044	1,044	1,038	1,048	1,051	1,047	1,044	1,044	1,047	1,048
제주/동부	298	685	693	687	685	679	682	681	676	678	678	681	685
제주/읍면	236	481	486	475	475	475	479	483	483	484	483	483	481

〈표 3-7〉 〈경활〉 2010년 연령 및 월별 응답 가구원수 (단위: 명)

연령	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전체	41,410	41,297	40,672	40,528	40,353	40,357	40,419	40,347	40,027	40,027	40,009	40,091
15-19세	3,578	3,590	3,317	3,331	3,363	<u>3,466</u>	<u>3,579</u>	<u>3,597</u>	3,504	3,549	<u>3,609</u>	<u>3,705</u>
20-24세	2,089	2,072	1,962	1,942	1,920	<u>1,959</u>	<u>1,997</u>	<u>1,999</u>	1,909	1,905	<u>1,922</u>	<u>1,966</u>
25-29세	3,003	2,975	2,910	2,898	2,840	2,827	2,816	2,799	2,761	2,751	2,706	2,699
30-34세	3,118	3,116	3,089	3,069	3,048	3,022	3,030	3,025	3,003	2,996	2,989	2,972
35-39세	4,055	4,041	4,014	3,991	3,966	3,948	3,942	3,927	3,910	3,918	3,909	3,900
40-44세	4,258	4,238	4,216	4,205	4,194	4,183	4,177	4,159	4,144	4,143	4,140	4,136
45-49세	4,518	4,513	4,483	4,467	4,460	4,447	4,429	4,423	4,402	4,400	4,398	4,391
50-54세	4,187	4,183	4,161	4,145	4,141	4,129	4,114	4,112	4,091	4,093	4,091	4,089
55-59세	3,021	3,009	2,996	2,985	2,972	2,966	2,964	2,967	2,960	2,952	2,949	2,949
60-64세	2,706	2,706	2,697	2,695	2,687	2,677	2,670	2,657	2,660	2,664	2,667	2,665
65-69세	2,286	2,286	2,282	2,279	2,261	2,255	2,249	2,246	2,247	2,240	2,235	2,234
70-74세	2,031	2,019	2,017	2,010	2,005	1,998	1,991	1,988	1,990	1,982	1,980	1,978
75-79세	1,387	1,387	1,379	1,370	1,366	1,359	1,355	1,349	1,346	1,342	1,331	1,326
80세 이상	1,173	1,162	1,149	1,141	1,130	1,121	1,106	1,099	1,100	1,092	1,083	1,081

아울러 2010년 1월 표본 가구 중 전출입 등에 의한 변동이 없었던 17,802가구의 44,592명의 가구원을 대상으로 총 응답 횟수를 분석한 결과는 <표 3-8>과 같다. 분석대상 응답자 중 2010년에 12개월 동안 모두 응답한 가구원은 37,365명(83.8%)으로 나타났다. 이런 응답률이 만족할만한 수준의 응답률인지 여부에 대해서는 논의가 필요할 것으로 보인다.

앞에서도 언급한 것처럼 표본 가구원 변동이 연령대별로 차이가 있을 수 있으며, 특히 청년층에서 이런 표본 구성원 변동이 심하게 발생할 소지가 높기 때문에 전체 응답자를 연령대별로 구분해 연간 응답횟수를 분석하였다. 전체적으로 12개월 모두 응답한 가구원 비율은 83.8%수준이었던 것에 비해, 연령대별로 보면 15~29세의 경우 58.9%, 20~24세의 경우 54.8%, 25~29세의 경우 70.9%로 전반적으로 청년층의 경우 12개월 지속적으로 응답하는 조사대상자 비율이 매우 낮은 것을 나타났다. 반면에 60~74세의 경우 12개월 동안 지속적으로 응답한 조사대상자의 비율이 93% 이상으로 대조적이다. 참고로, 앞에서도 언급한 것처럼 가구원 기준 응답패턴 분석에 사용된 경제활동인구조사의 현행 자료에는 동일 표본 가구 내에서 가구원이 전출입할 경우 동일인인 경우에도 가구원 코드가 변경되므로 일부 집계된 응답횟수가 정확하지 않을 수도 있다는 점을 다시 한 번 밝혀둔다. 그럼에도 불구하고 현재 표본 자료

분석결과 특별히 청년층에서 나타나는 표본 가구원 변동이 경제활동인구나 가계동향 관련 다양한 통계 작성에 있어서 어떤 영향을 줄 수 있는지에 대해서는 향후 추가적인 검토가 필요해 보인다.

〈표 3-8〉 〈경찰〉 2010년 응답횟수 및 연령별 응답자수

응답 횟수	전체	만 연령													
		15- 19세	20- 24세	25- 29세	30- 34세	35- 39세	40- 44세	45- 49세	50- 54세	55- 59세	60- 64세	65- 69세	70- 74세	75- 79세	80세 이상
전체(명)	44,592	4,690	2,750	3,372	3,330	4,203	4,380	4,633	4,301	3,104	2,771	2,340	2,070	1,418	1,230
12회 (명)	37,365	2,763	1,507	2,392	2,790	3,756	3,998	4,263	3,963	2,866	2,596	2,185	1,951	1,296	1,039
(%)	<u>83.8</u>	<u>58.9</u>	<u>54.8</u>	<u>70.9</u>	83.8	89.4	91.3	92.0	92.1	92.3	<u>93.7</u>	<u>93.4</u>	<u>94.3</u>	91.4	84.5
11회 (명)	552	122	78	78	47	40	46	42	38	11	10	15	6	10	9
(%)	1.2	2.6	2.8	2.3	1.4	1.0	1.1	0.9	0.9	0.4	0.4	0.6	0.3	0.7	0.7
10회 (명)	526	121	58	84	47	35	34	27	29	26	11	16	6	18	14
(%)	1.2	2.6	2.1	2.5	1.4	0.8	0.8	0.6	0.7	0.8	0.4	0.7	0.3	1.3	1.1
9회 (명)	432	100	50	67	28	28	28	20	31	18	18	13	8	11	12
(%)	1.0	2.1	1.8	2.0	0.8	0.7	0.6	0.4	0.7	0.6	0.7	0.6	0.4	0.8	1.0
8회 (명)	527	145	74	65	46	33	36	38	27	19	13	4	10	7	10
(%)	1.2	3.1	2.7	1.9	1.4	0.8	0.8	0.8	0.6	0.6	0.5	0.2	0.5	0.5	0.8
7회 (명)	504	110	80	76	33	31	30	34	27	23	20	10	4	10	16
(%)	1.1	2.4	2.9	2.3	1.0	0.7	0.7	0.7	0.6	0.7	0.7	0.4	0.2	0.7	1.3
6회 (명)	503	103	81	70	46	40	23	34	23	9	20	13	11	10	20
(%)	1.1	2.2	3.0	2.1	1.4	1.0	0.5	0.7	0.5	0.3	0.7	0.6	0.5	0.7	1.6
5회 (명)	585	118	87	82	56	41	27	41	32	22	18	16	13	13	19
(%)	1.3	2.5	3.2	2.4	1.7	1.0	0.6	0.9	0.7	0.7	0.7	0.7	0.6	0.9	1.5
4회 (명)	638	147	108	113	62	40	29	17	20	32	12	20	13	8	17
(%)	1.4	3.1	3.9	3.4	1.9	1.0	0.7	0.4	0.5	1.0	0.4	0.9	0.6	0.6	1.4
3회 (명)	732	193	140	84	48	56	33	40	30	24	21	18	12	11	22
(%)	1.6	4.1	5.1	2.5	1.4	1.3	0.8	0.9	0.7	0.8	0.8	0.8	0.6	0.8	1.8
2회 (명)	1274	507	258	133	84	50	37	47	39	24	20	14	16	16	29
(%)	2.9	10.8	9.4	3.9	2.5	1.2	0.8	1.0	0.9	0.8	0.7	0.6	0.8	1.1	2.4
1회 (명)	954	261	229	128	43	53	59	30	42	30	12	16	20	8	23
(%)	2.1	5.6	8.3	3.8	1.3	1.3	1.4	0.7	1.0	1.0	0.4	0.7	1.0	0.6	1.9

(2) 가계동향조사 응답 현황

경제활동인구조사와 마찬가지로 가계동향조사에서 표본 가구 및 가구원 응답패턴을 살펴보기 위해 연동표본으로 인한 조사구 변동이 발생하지 않는 2010년도의 4월, 8월, 12월 응답가구 현황을 정리해 본 결과는 <표 3-9>와 같다. 2010년 가계동향조사 대상인 9,990가구 중에서 연동이 없는 3개월 동안 평균 7,064가구(70.7%)가 응답한 것으로 나타나 경제활동인구조사에 비해 조사 참여율이 상당히 떨어진다는 것을 볼 수 있다.

<표 3-9> <가계> 층별 응답 현황

층	표본	가구		
		4월	8월	12월
전국	9,990	7,069	7,081	7,041
서울	1,260	840	854	862
부산	720	507	490	483
대구	540	417	408	418
인천	540	395	380	391
광주	540	359	374	392
대전	540	407	396	378
울산	360	257	258	257
경기/동부	720	534	543	528
경기/읍면	360	245	242	247
강원/동부	360	264	259	247
강원/읍면	270	198	209	204
충북/동부	270	198	202	193
충북/읍면	270	169	172	171
충남/동부	360	246	258	259
충남/읍면	270	195	201	196
전북/동부	360	264	263	253
전북/읍면	180	115	119	110
전남/동부	360	267	259	265
전남/읍면	180	122	122	116
경북/동부	360	273	276	279
경북/읍면	270	177	177	175
경남/동부	360	251	245	246
경남/읍면	270	191	194	193
제주/동부	180	124	124	122
제주/읍면	90	54	56	56

가계동향조사에서 조사대상 가구 중 지속적으로 응답한 가구 비율을 분석하기 위해 매년 연동에 의해 교체되는 조사구를 제외하면, 2008~2010년 연도별 조사대상 가구 규모는 <표 3-10>과 같으며, 연도별로 약간의 차이는 있지만 연동 교체 대상 가구를 제외하면 매년 5,800여 가구가 조사대상 가구인 것으로 나타났다. 한편 가계동향조사에서 12개월 가구 응답률을 산출하기 위해, 이들 5,800여 가구 중 1~11월에 표본에서 전출된 가구와 2~12월에 표본에 추가로 유입된 가구를 제외한 결과 응답률 분석대상이 되는 1월 응답 가구는 2008년에는 3,973가구, 2009년과 2010년은 각각 4,082가구, 4,166가구인 것으로 집계되었다. 실제 이들 가구는 원칙적으로 매월 조사에 응답해야하지만, 이 중 12개월 모두 응답한 가구는 2008년 89.7%, 2009년과 2010년은 각각 90.2%, 91.5%인 것으로 나타났다.

<표 3-10> <가계> 응답 가구 현황

연도 (제외한 연동그룹)	가구(가구)			
	연간	1월 기준	12회 응답	응답률(%)
2010년 (7~9)	5,825	4,166	3,810	91.5
2009년 (4~6)	5,852	4,082	3,682	90.2
2008년 (1~3)	5,873	3,973	3,562	89.7

여기서도 경제활동인구조사와 마찬가지로 동일 가구 전출입 외에 표본에 이탈/전입하는 가구가 많아 정확한 분석에 한계가 있으며, 동일 거처에 가구가 추가되는 경우도 많았다. 가계동향조사에서 표본 가구 이탈/전입 가구 현황을 정리하면 <표 3-11>과 같다. 경제활동인구조사와 비교해 보면 전출 및 전입 이외의 사유로 표본에 유입되거나 유출되는 가구 비율이 상대적으로 높은 것으로 나타났으며, 향후 가계동향조사의 안정성을 유지하기 위해서는 표본 가구의 유입 및 유출 사유를 지속적으로 관리하는 것이 필요해 보인다.

<표 3-11> <가계> 가구 이탈/전입 현황

연도 (제외한 연동그룹)	1-11월 마지막 응답						2-12월 처음 응답	전입	전입외	조사구 변동	가구 추가	기타
		전출	전출외	불응	제외	기타						
2010년 (7~9)	918	645	273	238	0	35	1,003	715	288	3	31	254
2009년 (4~6)	910	648	262	234	15	13	1,083	692	391	26	313	52
2008년 (1~3)	895	667	228	138	90	0	1,291	634	657	38	617	2

앞에서 언급한 것처럼 연동 교체 대상과 1~11월에 전출한 가구, 2~12월에 전입하거나 기타 사유로 2월 이후에 추가로 표본에 유입된 가구를 제외하면, 2010년의 가계동향조사 월별 가구 응답률을 산출하기 위한 분석대상은 1월에 응답한 4,166가구이다. 이들 2010년 가계동향조사 응답패턴 분석대상 가구를 기준으로 각 층에 대한 월별 응답가구수를 분석한 결과는 <표 3-12>와 같다. 매월 조사가 진행됨에 따라 응답가구수가 점점 감소하는 경향을 보인다는 것을 알 수 있으며, 2010년 1월을 기준으로 12월에는 응답률이 95.3%까지 감소하는 것으로 나타났다. 여기서도 무응답 이외에 표본에서 이탈한 가구의 경우 정확한 사유를 판단하기 어렵기 때문에 응답률 계산에 반영하지 않았다는 점을 참고하기 바란다.

〈표 3-12〉 〈가계〉 2010년(1월 기준) 층 및 월별 응답 가구수 (단위: 가구)

[illegible]

		합계	12회	11회	10회	9회	8회	7회	6회	5회	4회	3회	2회	1회
전체(가구)		4,166	3,810 91.5	122	48	27	27	28	16	15	14	12	23	24
서울	가구 (%)	489	439 89.8	14 2.9	7 1.4	3 0.6	3 0.6	6 1.2	3 0.6	1 0.2	1 0.2	4 0.8	2 0.4	6 1.2
부산	가구 (%)	310	274 88.4	11 3.6	4 1.3	3 1.0	2 0.7	4 1.3	2 0.7	0 0.0	4 1.3	1 0.3	3 1.0	2 0.7
대구	가구 (%)	250	231 92.4	10 4.0	2 0.8	0 0.0	1 0.4	0 0.0	2 0.8	1 0.4	2 0.8	0 0.0	1 0.4	0 0.0
인천	가구 (%)	229	206 90.0	6 2.6	5 2.2	2 0.9	2 0.9	1 0.4	2 0.9	1 0.4	0 0.0	0 0.0	1 0.4	3 1.3
광주	가구 (%)	222	202 91.0	5 2.3	2 0.9	1 0.5	1 0.5	1 0.5	0 0.0	3 1.4	1 0.5	0 0.0	5 2.3	1 0.5
대전	가구 (%)	231	213 92.2	8 3.5	0 0.0	3 1.3	3 1.3	1 0.4	0 0.0	2 0.9	1 0.4	0 0.0	0 0.0	0 0.0
울산	가구 (%)	148	135 91.2	2 1.4	1 0.7	0 0.0	1 0.7	2 1.4	0 0.0	2 1.4	0 0.0	2 1.4	2 1.4	1 0.7
경기/동부	가구 (%)	296	274 92.6	4 1.4	6 2.0	2 0.7	1 0.3	2 0.7	0 0.0	3 1.0	0 0.0	0 0.0	1 0.3	3 1.0
경기/읍면	가구 (%)	144	132 91.7	5 3.5	2 1.4	0 0.0	0 0.0	2 1.4	1 0.7	0 0.0	0 0.0	0 0.0	1 0.7	1 0.7
강원/동부	가구 (%)	156	144 92.3	6 3.9	2 1.3	0 0.0	2 1.3	1 0.6	0 0.0	0 0.0	0 0.0	0 0.0	1 0.6	0 0.0
강원/읍면	가구 (%)	132	123 93.2	4 3.0	2 1.5	0 0.0	0 0.0	1 0.8	1 0.8	0 0.0	0 0.0	0 0.0	0 0.0	1 0.8
충북/동부	가구 (%)	114	103 90.4	3 2.6	0 0.0	2 1.8	1 0.9	1 0.9	1 0.9	0 0.0	1 0.9	0 0.0	0 0.0	2 1.8
충북/읍면	가구 (%)	106	93 87.7	6 5.7	4 3.8	1 0.9	1 0.9	0 0.0	0 0.0	0 0.0	1 0.9	0 0.0	0 0.0	0 0.0
충남/동부	가구 (%)	139	126 90.7	5 3.6	2 1.4	1 0.7	1 0.7	0 0.0	1 0.7	0 0.0	0 0.0	0 0.0	1 0.7	2 1.4
충남/읍면	가구 (%)	118	113 95.8	2 1.7	1 0.9	1 0.9	0 0.0	1 0.9	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
전북/동부	가구 (%)	153	145 94.8	5 3.3	1 0.7	1 0.7	0 0.0	1 0.7	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
전북/읍면	가구 (%)	69	63 91.3	1 1.5	1 1.5	1 1.5	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	2 2.9	1 1.5	0 0.0
전남/동부	가구 (%)	154	146 94.8	2 1.3	0 0.0	1 0.7	3 2.0	0 0.0	0 0.0	0 0.0	1 0.7	0 0.0	1 0.7	0 0.0
전남/읍면	가구 (%)	71	64 90.1	4 5.6	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	1 1.4	2 2.8
경북/동부	가구 (%)	163	151 92.6	5 3.1	1 0.6	1 0.6	1 0.6	2 1.2	1 0.6	0 0.0	1 0.6	0 0.0	0 0.0	0 0.0
경북/읍면	가구 (%)	109	96 88.1	4 3.7	1 0.9	0 0.0	3 2.8	0 0.0	0 0.0	1 0.9	1 0.9	2 1.8	1 0.9	0 0.0
경남/동부	가구 (%)	149	134 89.9	5 3.4	4 2.7	3 2.0	0 0.0	0 0.0	1 0.7	0 0.0	0 0.0	1 0.7	1 0.7	0 0.0
경남/읍면	가구 (%)	110	106 96.4	1 0.9	0 0.0	1 0.9	1 0.9	1 0.9	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0
제주/동부	가구 (%)	71	64 90.1	4 5.6	0 0.0	0 0.0	0 0.0	1 1.4	1 1.4	1 1.4	0 0.0	0 0.0	0 0.0	0 0.0
제주/읍면	가구 (%)	33	33 100.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0	0 0.0

2) 집락 효과 분석

경제활동인구조사와 가계동향조사는 시도 및 읍면/동부 구분에 따라 구성된 25개 층별로 조사구를 추출하고 각 조사구에서 경제활동인구조사는 4개 구역의 20가구, 가계동향조사는 2개 구역의 10가구를 대상으로 조사를 실시한다. 이와 같은 복합표본설계에서는 층화 및 집락 추출에 따른 설계효과를 분석함으로써 향후 표본설계 개선을 위한 시사점을 찾을 수 있다. 특히 조사구 사용에 따른 집락효과를 분석해 봄으로써 현재 경활 및 가계동향조사에서 사용하는 조사구 내 표본 가구 규모의 적절성을 검토해 볼 수 있을 것이다. 물론 이런 집락효과는 어떤 변수를 분석대상으로 하느냐에 따라 차이가 있을 수 있다는 점을 염두에 두어야 한다. 따라서 본 연구에서는 경제활동인구조사의 경우 주요 변수에 해당하는 경활인구, 취업인구, 실업인구, 비경활인구, 취업시간을 분석대상으로 했으며 가계동향조사의 경우 가구의 소득, 경상소득, 근로소득, 가계지출, 소비지출 등 대부분의 주요 변수에 대한 집락효과 분석을 수행하고 그 결과를 종합적으로 검토한 것이다.

본 연구에서는 경제활동인구조사 및 가계동향조사에서 조사구를 집락으로 사용함으로써 발생하는 설계효과를 파악하기 위하여 SAS의 Surveymeans Procedure를 이용해 조사구를 집락으로 설정하는 경우와 조사구를 집락으로 설정하지 않는 경우 산출되는 분산추정량의 값을 비교하는 방식으로 집락효과를 분석하였다. 따라서 $var_{cluster}(\hat{\theta})$ 와 $var_{srs}(\hat{\theta})$ 를 조사구(집락) 단위로 가구를 추출했다고 가정하는 경우와 단순확률추출로 가구를 추출했다고 가정하는 경우 얻어지는 분산 추정량이라고 하면 집락효과는 다음과 같이 정의한 것이다.

$$\text{추정량의 집락효과} = \frac{Var_{du}(\hat{\theta})}{Var_{srs}(\hat{\theta})}$$

경제활동인구조사와 가계동향조사 모두 2010년 12개월 동안의 월별 자료를 대상으로 25개 층별로 집락효과를 분석하였다.

(1) 경제활동인구조사의 집락 효과

경제활동인구조사의 집락효과를 분석하기 위해 경활인구 총계, 취업인구 총계, 실업인구 총계, 비경활인구 총계 및 취업시간 평균에 대한 추정량의 분산을 SAS의 Surveymeans Procedure 이용해 추정한 후 그 결과를 이용해 산출하였다. 참고로, 비경활(d=1)인 경우 취업시간을 결측값으로 처리하였다 (괄호 안의 코드는 경제활동

인구조사 상의 변수 코드를 그대로 인용한 것임).

한편, 경활인구를 기준으로 한 집락효과는 <표 3-14>와 같으며, 집락효과는 2.78~6.71이며 1-12월의 전국 12개월 평균은 5.27이지만 월별로 집락효과가 상당히 차이가 있는 것으로 나타났다. 특히 조사구 연동 교체가 없는 4, 8, 12월은 그 밖의 다른 월과는 집락효과가 큰 차이가 있는데 이는 조사구 교체가 발생하는 월의 경우 조사구별로 표본 가구수가 달라지고 전체적으로 조사구 수가 늘어남에 따라 교체가 없는 달과는 상당히 다른 양상을 보이게 된다는 점에 유의할 필요가 있다.

취업인구에 대한 집락효과 분석결과는 <표 3-15>와 같다. 취업인구의 경우 집락효과는 2.59~6.32으로 경활인구와 큰 차이가 없으며 전국 12개월 평균은 4.90인 것으로 나타났다. 여기서도 월별로 집락효과가 큰 차이가 있으며 특히 조사구 연동 교체가 없는 4, 8, 12월은 낮은 것을 볼 수 있다.

<표 3-16>과 같이 실업인구의 집락효과는 1.20~1.60이며 12월 전국 평균은 1.36인 것으로 나타났다. 전국 기준으로는 월별 차이가 크지 않으며 경활인구나 취업인구와는 달리 집락효과가 크지 않다는 것을 알 수 있다.

한편 비경활인구의 집락효과는 <표 3-17>과 같이 전국 12월 평균 집락효과는 3.84인 것으로 나타났으며, 경활인구나 취업인구처럼 조사구의 연동 교체가 없는 4, 8, 12월의 경우 집락효과는 2.66~2.98인 반면, 이 들 3개월을 뺀 나머지 월에는 집락효과가 3.81~4.66으로 차이가 크다.

취업시간의 경우 집락효과는 <표 3-18>와 같이 집락효과가 1.52~1.92 수준이었으며, 전국 12월 평균은 1.62인 것으로 나타났다.

한편 집락효과 분석 결과 경제활동인구조사의 경우 변수별로 집락추출에 따른 설계효과에 있어서 상당히 큰 차이를 보였으며, 동일한 변수에 대해서도 월별 및 층별로 차이가 있다는 점을 참고하기 바란다. 아울러 여기 제시된 집락효과는 가구 내 가구원을 별개의 조사단위로 간주해 산출된 것이기 때문에 실제 조사구 내 가구를 집락으로 추출함으로써 발생하는 집락효과를 정확히 설명한다고 볼 수 없다. 집락 내 가구 추출에 따른 집락효과를 정확히 분석하기 위해서는 우선 각 가구 내의 가구원 자료를 종합해 가구 특성을 대표하는 변수로 전환한 후 분석하는 것이 필요하며, 2절의 '최적 표본 가구수 분석' 과정에서는 이런 방식으로 집락효과를 분석해 조사구 내 최적 가구수를 산출하였다는 점을 참고하기 바란다.

〈표 3-14〉 〈경찰〉 층 및 월별 경찰인구의 집락효과

층	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	6.19	5.39	6.09	<u>2.78</u>	6.71	5.87	6.55	<u>2.89</u>	6.31	5.41	6.20	<u>2.89</u>
서울	6.18	5.52	6.53	2.75	6.34	5.60	6.31	2.48	6.48	5.28	6.52	2.96
부산	6.11	5.48	5.90	2.96	6.70	6.09	6.70	3.04	5.86	5.22	5.82	2.57
대구	6.25	5.51	6.67	3.51	7.08	6.79	6.85	3.82	7.31	5.70	6.40	3.61
인천	6.68	5.70	6.49	2.56	7.29	6.77	7.69	3.11	7.10	6.02	6.79	3.94
광주	6.13	5.34	6.29	3.36	6.39	5.71	6.45	3.35	6.43	5.36	6.25	3.63
대전	6.01	5.16	5.76	2.76	7.10	6.27	6.70	2.91	5.64	4.87	5.44	2.66
울산	5.92	5.57	5.88	2.04	5.07	4.36	5.14	2.32	5.77	5.27	6.07	2.54
경기/동부	5.66	4.86	5.26	2.08	6.51	5.53	6.17	2.54	5.69	4.98	5.52	2.02
경기/읍면	6.42	6.36	7.18	3.43	7.38	6.12	6.45	3.21	7.03	5.93	7.39	3.73
강원/동부	5.77	5.40	5.52	3.22	5.87	4.92	5.36	2.72	5.83	4.74	4.55	2.54
강원/읍면	8.19	8.22	6.48	3.25	6.98	6.08	7.38	3.34	7.73	6.06	5.28	2.44
충북/동부	6.14	5.52	6.65	3.63	5.96	5.12	5.74	3.09	7.24	6.69	6.71	2.64
충북/읍면	9.20	8.98	8.51	6.17	10.26	9.06	8.45	4.87	10.34	9.98	9.87	5.44
충남/동부	5.46	5.66	6.97	2.78	6.32	4.95	5.83	1.97	5.80	5.18	5.92	2.16
충남/읍면	6.62	5.64	5.91	3.54	7.87	6.70	8.36	4.10	8.51	7.53	7.58	3.32
전북/동부	4.14	3.59	4.67	2.44	5.01	4.03	4.36	2.04	4.40	3.59	4.26	2.30
전북/읍면	8.16	7.53	8.66	4.43	7.35	5.66	7.15	3.06	6.11	6.90	8.19	4.71
전남/동부	6.74	6.10	6.21	3.00	6.66	5.97	6.73	3.55	6.96	6.76	7.16	3.36
전남/읍면	7.35	5.39	6.64	4.65	8.33	7.51	8.01	5.02	8.31	8.13	7.91	5.25
경북/동부	6.53	5.56	6.70	2.65	6.63	5.42	7.21	3.19	6.24	5.08	5.93	3.41
경북/읍면	9.95	8.04	8.49	5.28	7.48	7.35	8.17	4.41	7.25	6.14	6.83	5.19
경남/동부	7.13	5.75	7.09	4.17	8.61	8.39	9.30	5.16	7.85	6.69	7.72	4.84
경남/읍면	6.80	5.23	6.09	3.57	6.81	4.96	5.53	2.21	5.63	5.95	7.85	4.48
제주/동부	7.38	4.24	5.87	2.20	6.56	5.79	7.26	2.64	7.15	5.98	6.19	3.24
제주/읍면	6.41	4.77	6.84	2.41	7.64	7.08	9.19	3.17	6.13	3.62	5.82	2.24

〈표 3-15〉 〈경찰〉 총 및 월별 취업인구의 집락효과

총	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	5.61	4.93	5.57	<u>2.59</u>	6.32	5.35	6.06	<u>2.75</u>	5.91	5.10	5.88	<u>2.74</u>
서울	5.52	5.01	5.75	2.44	5.83	5.11	5.79	2.34	6.02	4.95	5.99	2.76
부산	5.55	5.15	5.68	2.84	6.39	5.67	6.19	2.98	5.42	4.97	5.66	2.64
대구	5.71	5.34	6.20	3.30	6.73	6.35	6.39	3.81	6.91	5.60	6.10	3.61
인천	6.19	5.22	5.97	2.47	6.86	6.30	6.82	2.87	6.50	5.79	6.56	3.85
광주	5.80	5.19	6.12	3.56	6.28	5.40	6.25	3.58	6.32	5.22	6.06	3.49
대전	5.66	4.94	5.51	2.84	6.77	6.14	6.33	3.05	5.53	4.68	5.14	2.56
울산	5.48	5.23	5.62	2.16	5.00	4.24	4.84	2.14	5.31	4.89	5.78	2.24
경기/동부	5.01	4.27	4.72	1.88	6.11	4.77	5.60	2.32	5.23	4.60	5.26	1.86
경기/읍면	5.92	5.75	6.47	3.17	7.09	5.92	6.06	2.97	6.69	5.70	6.88	3.49
강원/동부	5.66	5.07	5.47	3.17	5.65	4.58	5.44	2.79	5.71	4.41	4.33	2.49
강원/읍면	7.71	8.01	6.52	3.23	6.86	6.13	7.42	3.51	7.59	6.01	5.21	2.40
충북/동부	5.36	4.93	6.14	3.33	5.60	5.00	5.64	3.03	7.11	6.41	6.44	2.82
충북/읍면	8.86	8.74	8.59	6.14	10.26	9.10	8.30	4.87	10.04	9.72	9.53	5.19
충남/동부	4.60	5.11	6.31	2.64	5.90	4.66	5.43	1.85	5.66	4.83	5.32	1.97
충남/읍면	5.92	5.03	5.55	3.36	7.50	6.38	7.76	3.87	8.07	7.23	7.37	3.31
전북/동부	3.99	3.68	4.39	2.41	4.84	3.82	4.19	1.97	4.29	3.36	4.22	2.08
전북/읍면	8.15	7.68	8.69	4.32	7.17	5.50	7.15	3.13	6.10	6.43	7.98	4.46
전남/동부	6.23	5.84	6.07	3.19	6.64	5.94	6.54	3.76	6.71	6.40	6.95	3.52
전남/읍면	7.17	5.02	6.38	4.53	7.92	7.21	7.68	5.02	8.18	8.05	8.00	5.21
경북/동부	6.14	5.25	6.10	2.54	6.39	5.21	6.84	2.99	5.75	4.66	5.65	3.44
경북/읍면	9.26	7.68	8.11	5.06	7.35	7.09	8.00	4.39	7.01	5.92	6.63	4.77
경남/동부	6.77	5.67	7.01	4.29	8.12	7.84	8.81	5.18	7.69	6.45	7.40	4.50
경남/읍면	6.11	5.07	5.82	3.24	6.49	4.59	5.14	2.14	5.56	5.88	7.65	4.31
제주/동부	7.10	4.26	5.78	1.99	6.05	5.48	6.82	2.38	6.70	5.75	5.97	3.06
제주/읍면	6.20	4.41	6.87	2.25	7.29	6.93	9.09	3.07	6.10	3.68	5.56	2.20

〈표 3-16〉 〈경찰〉 총 및 월별 실업인구의 집락효과

총	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	1.60	1.42	1.38	1.27	1.34	1.43	1.46	1.49	1.22	1.20	1.27	1.24
서울	1.75	1.69	1.78	1.68	1.77	1.74	1.53	1.79	1.33	1.30	1.32	1.21
부산	1.18	1.13	1.12	1.03	1.09	1.09	1.37	1.21	1.39	1.40	1.31	1.56
대구	1.56	1.54	1.61	1.18	1.16	1.31	1.43	1.40	1.44	1.08	1.03	0.92
인천	1.57	1.20	1.36	1.14	1.28	1.19	1.33	1.19	1.22	1.22	1.44	1.39
광주	1.12	1.34	1.52	1.35	1.05	1.32	1.99	1.34	1.28	1.24	1.07	1.43
대전	1.59	1.11	1.45	1.37	1.31	1.21	1.22	1.21	1.12	1.31	1.28	1.71
울산	1.04	1.36	1.44	1.24	1.52	1.21	1.47	1.31	1.32	1.19	1.50	1.18
경기/동부	1.64	1.28	1.11	1.01	1.12	1.41	1.47	1.40	1.08	1.12	1.16	1.14
경기/읍면	1.35	1.25	1.07	1.02	1.10	0.92	1.15	1.21	1.06	1.01	1.10	0.96
강원/동부	1.27	1.52	1.19	0.84	0.95	0.99	1.39	1.76	1.29	1.19	1.34	1.37
강원/읍면	1.37	2.22	2.48	0.97	1.16	0.99	1.15	1.15	1.16	1.29	1.13	1.58
충북/동부	1.79	1.77	1.15	1.30	1.18	1.31	1.44	1.42	0.79	1.32	1.02	1.55
충북/읍면	1.42	1.18	1.00	0.94	0.90	0.85	1.01	1.35	1.68	0.97	0.95	1.11
충남/동부	1.29	1.15	1.24	1.37	1.40	1.03	1.31	1.38	1.21	1.56	1.89	1.85
충남/읍면	1.59	2.09	1.55	1.31	1.01	1.00	1.17	0.98	1.29	1.06	1.26	1.15
전북/동부	0.78	1.36	1.31	1.34	0.97	0.99	1.30	1.00	0.92	0.97	1.57	1.64
전북/읍면	1.42	1.06	1.53	0.97	1.46	0.99	1.21	0.94	0.92	1.20	1.19	1.01
전남/동부	1.37	1.21	1.62	1.42	1.20	1.71	1.34	1.02	0.88	0.81	0.82	1.57
전남/읍면	1.82	1.57	1.26	1.16	1.24	0.96	0.97	1.31	0.97	1.20	1.12	1.26
경북/동부	1.10	1.53	1.60	1.44	1.34	1.21	1.52	1.56	1.22	1.32	1.45	1.64
경북/읍면	1.86	1.19	1.23	0.95	1.05	1.14	1.07	1.81	1.40	1.17	0.91	0.88
경남/동부	1.31	1.23	1.11	1.07	1.31	1.43	1.90	1.66	1.48	1.19	1.99	1.87
경남/읍면	1.54	1.78	1.45	1.41	1.08	1.16	1.31	0.98	0.88	0.79	1.21	0.99
제주/동부	1.38	1.69	0.83	1.13	1.28	0.88	1.44	1.20	1.01	1.47	1.15	1.89
제주/읍면	1.00	1.29	1.57	0.89	0.89	0.96	1.22	1.22	1.54	1.14	1.35	1.09

〈표 3-17〉 〈경찰〉 층 및 월별 비경찰인구의 집락효과

층	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	4.66	4.00	4.18	<u>2.66</u>	4.23	3.81	4.18	<u>2.72</u>	4.36	3.91	4.42	<u>2.98</u>
서울	4.46	3.68	4.23	2.73	4.33	3.80	4.28	2.73	4.47	4.13	4.41	3.19
부산	5.10	4.57	4.76	2.20	4.84	4.03	4.44	2.54	4.28	3.90	3.89	2.17
대구	4.38	4.10	4.12	2.34	3.93	3.71	4.39	2.44	4.06	3.62	4.23	2.20
인천	3.61	3.12	3.50	2.33	3.51	3.02	3.33	2.30	4.15	3.35	3.52	2.22
광주	4.15	3.25	3.92	2.24	3.92	3.33	3.90	2.08	3.43	2.92	4.05	2.18
대전	4.50	3.75	4.06	2.26	4.08	3.82	4.48	2.58	4.00	3.32	4.19	2.22
울산	3.99	3.75	3.84	1.73	3.09	2.77	3.26	1.78	3.67	3.59	3.87	2.23
경기/동부	5.19	4.47	4.49	3.01	4.79	4.33	4.46	3.07	4.77	4.13	4.98	3.45
경기/읍면	5.16	4.17	3.86	2.39	3.15	2.54	3.13	2.33	3.95	3.71	3.75	3.00
강원/동부	3.76	3.84	4.38	2.56	4.34	3.78	4.12	2.21	3.89	3.32	3.72	2.16
강원/읍면	6.65	5.57	5.37	2.86	3.30	3.49	4.05	2.48	3.86	3.95	4.75	3.58
충북/동부	3.38	2.84	3.50	2.66	4.19	3.38	4.28	2.55	4.03	3.54	3.73	1.99
충북/읍면	4.02	3.03	2.97	2.37	3.02	2.97	2.90	2.08	4.04	3.40	3.44	2.48
충남/동부	4.26	4.13	3.76	2.16	3.35	3.42	3.99	2.70	3.62	3.13	3.70	2.02
충남/읍면	5.87	5.47	4.43	2.97	3.70	4.02	4.24	2.62	3.69	3.46	4.26	3.25
전북/동부	4.11	4.03	3.60	2.06	3.97	3.26	3.66	2.02	4.24	3.96	4.05	2.21
전북/읍면	3.75	3.21	3.43	2.09	2.36	2.27	2.87	2.20	3.23	3.09	3.51	1.87
전남/동부	3.78	3.67	3.97	2.21	3.27	3.16	3.64	2.21	3.68	2.63	3.52	2.03
전남/읍면	3.72	3.53	3.28	2.18	2.66	2.82	4.03	3.05	3.41	3.02	3.61	2.27
경북/동부	3.85	2.91	3.40	1.86	2.75	2.72	3.74	2.18	3.14	3.33	3.34	1.50
경북/읍면	4.29	3.20	3.95	3.16	4.39	4.52	4.40	3.20	3.69	3.60	3.84	2.79
경남/동부	4.06	3.92	3.68	2.27	3.37	2.99	3.72	2.40	3.88	3.75	4.42	3.23
경남/읍면	3.47	2.80	2.84	1.73	2.78	2.77	3.19	2.08	4.40	4.21	4.16	2.76
제주/동부	1.85	1.72	2.73	2.06	3.36	3.34	3.55	1.24	2.71	3.13	3.56	1.37
제주/읍면	4.27	4.42	4.43	2.81	3.92	4.21	4.02	2.34	2.72	2.90	2.58	2.54

〈표 3-18〉 〈경찰〉 층 및 월별 취업시간(1주)의 집락효과

층	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	1.73	1.92	1.64	1.61	1.56	1.55	1.57	1.69	1.55	1.52	1.47	1.63
서울	1.34	1.63	1.55	1.73	1.80	1.68	1.60	1.70	1.60	1.62	1.44	1.55
부산	1.60	2.52	1.43	1.39	1.28	1.31	1.66	1.60	1.26	1.10	1.24	1.34
대구	1.50	1.55	1.42	1.33	1.40	1.54	1.70	1.91	1.75	1.62	1.64	1.20
인천	1.30	1.26	1.22	1.25	1.01	1.13	1.13	1.08	1.32	1.19	1.09	1.21
광주	1.75	2.08	1.93	2.03	1.34	1.66	1.61	1.38	1.34	1.26	1.21	1.69
대전	0.93	1.19	1.18	1.03	1.09	0.98	1.36	1.01	0.89	1.06	0.93	1.15
울산	1.18	1.79	1.37	1.14	1.24	1.31	1.26	1.24	1.25	1.34	1.40	1.24
경기/동부	1.63	1.85	1.43	1.37	1.30	1.27	1.39	1.61	1.31	1.35	1.24	1.31
경기/읍면	2.07	1.60	1.70	1.54	1.68	1.65	1.66	1.38	1.35	1.20	1.90	2.54
강원/동부	1.09	1.31	0.87	0.74	1.15	0.67	1.42	1.70	1.32	1.11	1.30	1.44
강원/읍면	6.98	8.82	2.12	1.88	1.67	1.76	2.23	2.08	1.95	1.70	2.43	2.30
충북/동부	1.53	0.93	1.40	1.23	1.34	1.14	1.57	1.63	1.42	1.28	1.23	1.27
충북/읍면	3.30	2.20	2.51	1.71	2.55	2.85	2.50	3.67	3.15	2.67	2.27	3.59
충남/동부	1.56	1.44	1.57	1.51	1.02	0.99	1.33	1.59	0.94	1.55	1.53	1.58
충남/읍면	3.20	2.32	3.02	2.10	1.91	3.11	2.21	2.96	3.40	2.74	1.80	1.82
전북/동부	1.29	2.31	1.69	1.76	1.37	1.45	1.62	1.55	1.20	1.59	1.52	2.12
전북/읍면	5.42	5.02	4.62	3.97	2.68	2.85	2.63	2.62	1.73	1.39	2.32	3.89
전남/동부	1.99	2.63	2.23	1.56	1.84	1.56	1.46	1.29	1.39	1.38	1.23	1.50
전남/읍면	4.62	4.26	3.77	4.53	4.39	3.03	3.51	3.20	4.19	3.43	4.08	5.66
경북/동부	2.18	1.88	2.40	1.31	1.41	1.30	1.25	1.40	1.26	1.57	1.38	2.75
경북/읍면	4.80	3.14	3.43	2.43	2.59	4.03	2.32	3.50	2.89	2.69	3.07	3.44
경남/동부	1.49	1.95	1.56	1.45	1.34	1.26	1.16	1.12	1.51	1.81	1.68	1.89
경남/읍면	3.08	2.80	3.40	2.96	2.17	1.97	2.48	2.32	2.24	1.48	2.08	2.85
제주/동부	1.54	1.77	1.55	1.49	1.66	1.21	1.72	1.53	1.34	1.67	1.94	2.40
제주/읍면	2.37	2.82	2.37	2.33	2.32	2.01	2.09	1.18	2.13	2.02	2.41	3.82

※ 비경찰(d=1)인 경우, 취업시간(HSGJNO13_3)을 결측값으로 처리하여 분석

(2) 가계동향조사의 집락 효과

가계동향조사의 집락효과를 분석하기 위해 소득, 경상소득, 근로소득, 가계지출, 소비지출 평균 추정량에 대한 분산을 구하여 분석하였다.

소득을 분석대상 변수로 하는 경우 <표 3-19>와 같이, 집락효과는 전국 기준 1.29~2.25이며 전국 12개월 평균 집락효과는 1.97로 나타났다.

경상소득의 경우 <표 3-20>과 같이 집락효과는 전국 기준 1.98~2.40이며 12개월 평균은 2.17로 나타났다.

근로소득의 집락효과는 <표 3-21>과 같이 전국 기준 1.76~2.09이며 1-12월의 평균은 1.86으로 계산되었으며, 제주 읍면부 지역은 1보다 낮게 나타났다.

가계동향조사 지출 부문 중 주요 변수인 가계지출 자료를 기준으로 한 집락효과는 <표 3-22>와 같이 전국 기준으로 집락효과는 1.67~2.28이며 12개월의 전국 평균은 1.99로 나타났다.

한편 소비지출의 집락효과는 <표 3-23>와 같으며, 전국 기준 1.36~1.92이며 12개월 평균은 1.67로 나타났다.

〈표 3-19〉 <가계> 층 및 월별 소득의 집락효과

층	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	1.29	2.01	1.91	2.10	2.01	2.15	2.00	2.07	2.01	1.98	1.91	2.25
서울	2.14	2.04	1.77	1.78	1.93	2.22	1.84	2.07	1.85	1.87	1.93	2.19
부산	1.34	1.62	1.36	1.35	1.19	1.41	1.37	1.25	1.43	1.18	1.28	1.69
대구	2.26	2.23	2.33	2.15	2.14	2.23	2.29	2.12	2.29	1.69	2.01	1.85
인천	1.74	1.38	1.65	2.04	1.62	1.80	1.40	1.59	1.60	1.34	1.47	1.27
광주	2.15	2.06	1.88	2.00	2.02	1.73	1.74	1.97	2.17	1.65	2.11	2.16
대전	1.12	2.18	2.57	2.65	2.58	2.15	2.35	2.33	2.30	2.84	2.05	2.24
울산	1.47	1.74	1.60	2.07	1.72	2.42	2.62	2.53	2.48	1.84	1.36	2.75
경기/동부	1.23	2.21	2.17	2.42	2.55	2.29	2.19	2.30	2.56	2.28	2.09	2.55
경기/읍면	1.21	1.78	1.35	1.46	1.51	1.89	1.42	1.41	0.94	1.63	1.25	2.04
강원/동부	1.53	1.27	1.53	1.42	1.16	1.55	1.33	1.47	1.53	1.26	1.44	1.51
강원/읍면	2.28	2.59	2.32	2.12	1.70	1.54	1.62	1.88	1.89	1.84	1.43	1.50
충북/동부	2.07	1.88	2.17	2.42	1.60	1.60	2.29	2.09	2.74	1.66	1.49	2.52
충북/읍면	1.89	1.70	1.54	1.39	1.45	1.87	1.61	1.40	1.54	1.34	1.22	1.27
충남/동부	2.11	2.01	1.98	1.38	2.26	2.09	2.14	2.62	2.49	2.64	2.36	2.42
충남/읍면	1.48	1.79	1.35	1.92	2.41	2.61	3.03	2.76	3.00	3.12	1.79	2.74
전북/동부	1.12	1.59	1.28	1.91	1.78	1.31	1.61	1.54	1.52	1.46	1.62	1.24
전북/읍면	3.09	2.77	2.81	1.94	2.91	3.28	3.14	3.31	2.86	3.46	1.58	2.77
전남/동부	2.01	1.99	1.98	2.29	1.98	2.13	2.11	2.66	1.16	2.58	1.84	2.53
전남/읍면	1.78	1.33	2.01	1.62	1.16	1.48	1.11	2.15	1.43	2.17	1.51	1.81
경북/동부	1.85	2.03	2.21	2.58	2.44	2.71	2.61	2.65	2.65	2.02	2.68	2.72
경북/읍면	2.59	2.98	2.03	2.39	2.79	2.51	2.53	2.31	2.75	1.89	2.16	1.91
경남/동부	3.10	2.66	2.65	3.02	1.98	2.41	2.23	1.69	1.72	2.21	2.31	1.78
경남/읍면	1.60	3.06	1.81	2.92	2.14	1.99	1.94	1.49	1.89	1.71	1.31	2.10
제주/동부	1.38	1.72	1.58	1.54	1.41	1.54	2.10	1.97	2.05	1.70	1.42	1.85
제주/읍면	1.88	1.33	1.23	1.87	1.19	1.38	1.53	1.84	1.13	1.02	1.24	1.21

〈표 3-20〉 〈가계〉 총 및 월별 경상소득의 집락효과

총	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	2.40	2.14	1.98	2.26	2.20	2.19	2.06	2.13	2.11	2.17	2.09	2.28
서울	2.16	2.01	1.77	1.92	1.94	2.17	1.87	2.08	1.88	1.97	2.07	2.16
부산	1.51	1.58	1.37	1.40	1.43	1.37	1.37	1.45	1.49	1.42	1.47	1.63
대구	2.25	2.47	2.37	2.29	2.17	2.36	2.33	2.16	2.38	1.97	2.21	2.39
인천	2.15	2.35	2.05	2.11	1.68	1.77	1.57	1.51	1.67	1.62	1.42	1.26
광주	2.19	1.99	1.84	1.91	1.98	1.90	1.72	1.99	2.15	2.07	2.09	2.06
대전	2.76	2.85	2.62	2.62	2.71	2.33	2.28	2.38	2.37	2.51	2.03	2.27
울산	1.50	1.66	1.62	2.05	1.72	2.16	2.58	2.71	2.63	1.89	1.37	3.49
경기/동부	2.70	2.27	2.28	2.61	2.73	2.37	2.37	2.31	2.53	2.60	2.22	2.52
경기/읍면	1.84	1.75	1.31	1.48	1.56	1.90	1.36	1.69	1.77	1.64	1.41	2.18
강원/동부	1.46	1.19	1.37	1.24	1.26	1.58	1.46	1.59	1.59	1.42	1.48	2.01
강원/읍면	2.34	2.54	2.65	2.19	1.80	1.59	1.83	1.87	2.02	1.85	1.63	1.76
충북/동부	2.18	1.90	2.21	2.44	2.45	2.08	2.38	2.20	2.72	2.34	1.79	2.51
충북/읍면	1.92	1.69	1.54	1.40	1.60	1.82	1.63	1.41	1.57	1.58	1.15	1.26
충남/동부	2.33	2.25	1.90	2.51	2.47	2.05	2.75	2.58	2.70	2.71	2.37	2.40
충남/읍면	1.57	1.62	1.35	1.93	2.37	2.62	2.92	2.93	2.87	3.01	2.89	2.77
전북/동부	1.72	1.67	1.32	1.93	1.94	1.44	1.72	1.57	1.68	1.63	1.79	1.91
전북/읍면	3.09	2.75	2.86	2.57	2.96	3.29	3.12	3.31	3.27	3.18	3.03	2.78
전남/동부	2.01	1.97	2.15	2.40	2.03	2.44	2.16	2.72	1.16	2.58	2.50	2.50
전남/읍면	1.50	1.34	1.89	1.69	1.27	1.57	1.09	2.15	1.50	2.00	1.42	1.85
경북/동부	1.77	2.20	2.55	2.37	2.85	2.73	2.64	2.62	2.57	2.03	2.67	3.06
경북/읍면	2.69	2.89	2.38	2.31	2.39	2.33	2.60	2.39	2.76	1.77	2.22	1.81
경남/동부	3.08	2.77	2.69	3.15	2.38	2.65	2.14	1.90	2.27	2.19	2.25	1.71
경남/읍면	1.63	3.08	2.61	2.88	1.94	1.90	1.92	1.53	1.80	1.46	1.31	2.36
제주/동부	1.40	1.78	1.60	1.93	1.94	1.83	2.39	1.96	2.09	1.91	2.21	1.82
제주/읍면	2.03	1.28	1.19	1.24	1.04	1.02	1.37	1.31	1.16	1.39	1.23	1.12

〈표 3-21〉 〈가계〉 총 및 월별 근로소득의 집락효과

총	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	2.09	1.86	1.79	2.05	1.88	1.85	1.81	1.80	1.79	1.76	1.77	1.87
서울	1.59	1.69	1.53	1.63	1.62	1.80	1.60	1.73	1.58	1.62	1.74	1.74
부산	1.01	1.11	1.09	1.05	1.20	1.24	1.14	1.11	1.26	1.22	1.17	1.34
대구	1.71	1.93	1.82	1.67	1.46	1.75	1.69	1.48	1.51	1.34	1.56	1.83
인천	1.80	2.01	1.55	1.86	1.75	1.99	1.75	1.56	1.53	1.74	1.36	1.24
광주	1.68	1.56	1.39	1.35	1.37	1.32	1.19	1.40	1.65	1.59	1.51	1.45
대전	2.30	2.28	1.97	1.97	1.84	1.87	1.48	1.51	1.70	1.73	1.68	1.35
울산	1.56	1.73	2.20	2.34	1.96	2.36	2.87	2.70	2.83	2.21	1.83	3.39
경기/동부	2.42	1.96	2.15	2.47	2.35	2.02	2.15	2.04	2.09	2.03	1.98	1.99
경기/읍면	1.87	1.89	1.61	1.64	1.60	1.83	1.27	1.56	1.57	1.63	1.39	1.53
강원/동부	1.25	0.99	1.18	1.33	1.42	1.86	1.53	1.79	1.66	1.67	1.43	1.76
강원/읍면	2.70	3.06	3.24	2.54	2.42	2.18	2.28	1.90	2.13	2.19	2.11	1.65
충북/동부	1.96	1.85	1.71	1.86	1.67	1.41	1.62	1.67	2.08	1.89	1.80	2.22
충북/읍면	1.84	1.57	1.46	1.41	1.48	1.49	1.58	1.47	1.37	1.25	1.28	1.33
충남/동부	1.68	1.92	1.13	2.01	2.27	2.05	2.02	1.94	2.26	2.04	1.89	2.20
충남/읍면	1.46	1.64	1.70	1.66	1.62	2.15	2.11	2.47	2.56	2.59	2.14	2.23
전북/동부	1.18	1.29	1.25	1.72	1.48	1.42	1.31	1.12	1.30	1.37	1.52	1.47
전북/읍면	1.20	1.24	1.28	1.30	1.46	1.53	1.55	1.54	1.54	1.35	1.47	1.49
전남/동부	1.72	1.36	1.58	1.63	1.50	1.68	1.61	2.02	2.14	1.92	1.94	1.76
전남/읍면	1.92	2.44	2.09	1.79	1.56	1.60	1.74	1.57	1.57	1.29	1.74	1.19
경북/동부	1.94	2.05	2.63	2.47	2.52	2.59	2.37	2.80	2.67	2.63	2.40	3.17
경북/읍면	2.73	2.74	2.58	2.52	2.74	2.33	2.55	2.20	2.76	1.80	2.15	1.93
경남/동부	2.55	2.22	1.74	1.61	1.29	1.36	1.35	1.34	1.49	1.38	1.30	1.72
경남/읍면	1.86	3.32	3.26	3.46	2.34	2.56	2.11	1.55	1.74	1.65	1.53	2.49
제주/동부	1.75	2.24	2.07	2.14	2.27	2.11	1.94	1.89	1.85	1.59	1.75	1.36
제주/읍면	2.26	1.42	0.77	0.91	0.51	0.57	0.53	0.51	0.37	0.52	0.33	0.62

〈표 3-22〉 〈가계〉 층 및 월별 가계지출의 집락효과

층	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	2.28	2.19	1.94	1.79	2.18	1.95	1.97	1.86	2.19	1.67	2.00	1.88
서울	1.93	2.17	1.75	1.72	1.92	2.02	1.81	1.81	2.23	1.78	2.22	1.98
부산	1.53	1.23	1.76	1.47	1.79	1.60	1.98	1.30	1.31	1.48	1.26	1.57
대구	2.62	2.36	2.44	1.84	2.07	2.02	2.71	2.14	2.49	1.55	2.33	2.13
인천	1.82	1.93	2.14	2.27	2.23	1.66	1.99	2.11	2.09	1.30	1.85	1.35
광주	1.96	1.64	1.64	1.35	2.05	1.82	2.26	2.46	2.11	1.85	1.81	1.55
대전	2.43	1.86	2.13	2.69	2.14	1.36	2.07	2.37	2.62	2.01	2.38	2.14
울산	1.21	1.23	1.86	2.10	1.32	1.57	2.09	1.93	1.89	1.26	1.95	1.96
경기/동부	3.05	2.73	2.28	1.83	2.82	2.07	2.21	1.99	2.58	1.59	2.07	2.16
경기/읍면	1.65	1.30	1.32	1.32	1.96	1.53	1.49	1.28	0.81	1.90	1.30	1.30
강원/동부	1.64	2.26	1.93	1.76	1.50	1.32	1.26	1.38	1.62	1.15	1.61	1.30
강원/읍면	1.41	1.55	1.71	2.40	2.17	1.89	1.11	1.55	1.58	2.33	1.38	1.55
충북/동부	2.08	1.11	1.90	2.12	2.10	1.84	1.11	1.84	2.10	1.62	2.22	2.12
충북/읍면	1.19	1.19	1.09	1.81	1.42	1.43	1.23	0.91	1.48	1.11	1.08	1.18
충남/동부	2.50	2.51	2.04	1.05	2.47	2.46	2.43	2.29	1.56	2.53	1.32	2.23
충남/읍면	1.11	1.83	1.30	1.57	1.90	1.28	1.42	3.11	1.65	1.65	1.96	1.25
전북/동부	1.47	1.24	1.18	2.02	2.17	1.58	1.97	1.47	1.86	1.90	1.44	1.60
전북/읍면	3.22	2.35	2.80	3.20	1.93	3.02	3.22	3.00	1.66	2.85	1.07	2.63
전남/동부	2.32	1.30	1.93	2.51	1.96	1.80	1.73	2.47	1.96	2.16	1.93	2.28
전남/읍면	1.83	1.83	1.81	1.94	1.50	1.24	1.81	1.79	2.05	2.31	1.59	1.69
경북/동부	2.20	2.17	1.98	2.30	2.39	2.28	2.73	2.37	2.75	2.01	2.12	2.18
경북/읍면	2.31	1.80	1.47	2.74	2.61	2.45	2.37	2.08	2.23	2.44	2.23	1.56
경남/동부	2.46	2.36	2.27	2.49	1.85	2.65	2.01	1.54	1.82	2.49	2.36	1.37
경남/읍면	1.42	1.97	2.55	2.28	1.79	2.60	2.06	1.77	1.61	1.59	1.54	1.59
제주/동부	1.70	1.36	1.52	1.90	1.55	1.15	1.40	1.55	1.85	0.79	1.42	1.81
제주/읍면	1.72	1.45	1.13	1.07	0.72	1.38	1.10	0.95	1.35	0.30	1.04	0.76

〈표 3-23〉 〈가계〉 총 및 월별 소비지출의 집락효과

총	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	1.92	1.91	1.69	1.51	1.87	1.62	1.55	1.64	1.71	1.36	1.73	1.53
서울	1.73	2.06	1.51	1.48	1.73	1.68	1.49	1.57	1.90	1.54	1.98	1.75
부산	1.40	1.14	1.70	1.26	1.68	1.47	1.93	1.19	1.10	1.39	1.20	1.47
대구	2.40	2.04	2.31	1.68	1.79	1.69	2.26	2.14	2.23	1.35	2.21	1.95
인천	1.48	1.68	1.78	1.82	1.83	1.45	1.65	1.83	1.91	1.14	1.79	1.20
광주	1.66	1.42	1.66	1.26	1.83	1.58	2.09	2.29	1.89	1.57	1.60	1.19
대전	2.00	1.45	1.69	2.26	1.72	1.04	1.69	1.96	1.92	1.58	2.03	1.92
울산	1.20	1.11	1.82	1.96	1.30	1.34	1.57	1.71	1.25	1.03	1.65	1.57
경기/동부	2.56	2.28	1.91	1.48	2.28	1.69	1.59	1.73	1.84	1.21	1.73	1.59
경기/읍면	1.44	1.21	1.39	1.00	1.81	1.36	1.39	1.25	0.76	1.71	1.17	1.10
강원/동부	1.50	2.10	1.81	1.55	1.58	1.20	1.17	1.11	1.33	1.05	1.33	1.11
강원/읍면	1.29	1.26	1.29	1.67	1.49	1.21	1.07	1.31	1.55	1.93	1.30	1.43
충북/동부	1.92	1.06	1.75	1.83	1.99	1.60	0.99	1.64	1.67	1.18	1.93	1.71
충북/읍면	1.08	1.12	1.15	1.59	1.41	1.63	1.15	0.77	1.49	1.21	0.96	1.16
충남/동부	2.09	2.27	1.92	0.93	2.16	2.13	1.95	2.06	1.28	2.21	1.09	1.98
충남/읍면	1.04	1.64	1.66	1.62	2.19	0.97	1.23	3.07	1.22	1.44	1.66	1.13
전북/동부	1.54	1.24	1.20	1.78	2.02	1.49	1.64	1.36	1.82	1.74	1.31	1.31
전북/읍면	2.76	2.40	2.20	2.95	1.66	2.85	2.90	2.59	1.55	2.74	1.00	2.52
전남/동부	2.13	1.16	1.55	2.29	1.68	1.39	1.49	2.60	1.68	2.01	1.58	2.04
전남/읍면	1.61	1.78	1.31	1.70	1.37	0.99	1.82	1.56	1.91	2.15	1.39	1.63
경북/동부	2.03	1.93	1.70	1.94	1.85	1.77	2.27	2.12	2.28	1.73	1.77	1.69
경북/읍면	2.02	1.62	1.22	2.75	2.43	1.97	1.90	2.01	1.85	2.25	1.95	1.64
경남/동부	2.43	2.04	2.13	2.32	1.55	2.56	1.78	1.48	1.53	2.39	2.43	1.25
경남/읍면	1.23	1.93	2.21	1.93	1.56	2.41	1.78	1.53	1.22	1.29	1.32	1.35
제주/동부	1.76	1.44	1.67	1.70	1.45	1.10	1.31	1.50	1.96	0.78	1.34	1.64
제주/읍면	1.65	1.37	0.99	0.90	0.82	1.45	1.51	0.94	1.28	0.42	1.18	0.61

3) 조사구 정렬 변수 분석

현행 통계청의 가구부문 표본설계에서는 개략적으로 전체 조사구를 25개 층으로 나눈 후 층 내에서 거처유형, 광공업비율, 농림어업비율, 전월세비율, 행정구역+총조사번호 등을 기준으로 정렬 후 계통추출 한다. 경제활동인구조사를 포함한 통계청의 가구부문 표본설계에 있어서 지역구분에 따른 층화 후 정렬변수는 실제 내재적 층화(implicit stratification)를 통해 표본의 대표성을 제고한다는 차원에서 매우 중요한 역할을 하는 변수들이며, 어떤 변수들을 정렬 변수로 사용할 것인지에 따라 가구부문 표본설계의 효율성이 큰 영향을 주게 된다. 따라서 여기서는 현행 통계청 가구부문 표본설계에서 층화 후 정렬에 사용되는 분류기준 변수들의 타당성을 검증해 보기로 한다. 이를 위해 2005년 인구주택총조사에서 얻은 조사구별 정렬 기준 변수들과 실제 경제활동인구조사를 통해 조사되는 주요 변수 및 가계동향조사에서 조사되는 주요 변수와 각각 상관분석을 실시하였다.

본 연구에서는 우선 현재 경제활동인구조사 표본 조사구와 2005년 인구주택총조사 자료에서 조사구를 서로 연계시키는 작업을 통해 인구주택총조사와 경제활동인구조사 통합 자료를 먼저 구축한 후 상관분석을 하였다. 물론 인구주택총조사 자료는 2010년 경제활동인구조사나 가계동향조사 자료와 비교해 보면 5년이라는 시차가 있는 정보이지만, 그럼에도 불구하고 실제 표본설계에서는 이런 시차를 갖는 조사구 정보를 사용할 수밖에 없다는 측면을 고려하면, 본 연구에서 얻은 분석결과는 시사하는 바가 크다고 할 수 있다. 여기서는 조사구 정렬기준 변수와 상관분석 대상 변수로 경제활동인구조사의 경우 실업률과 경활비율을 사용하였고, 가계동향조사에서는 경상소득, 근로소득, 사업소득, 교육, 주거 및 수도광열, 보건변수 등을 대상으로 분석하였다.

(1) 경제활동인구조사 주요 변수와 정렬기준변수와의 상관분석

광공업비율, 농림어업비율, 전월세가구비율에 따른 조사구별 실업률과 경활비율과의 상관분석을 25개 층별로 실시하였는데, 상관분석 결과는 <표 3-24>와 같다. 참고로 <표 3-24>에서 진하게 표시된 것들은 상관계수가 통계적으로 유의하다는 것을 의미한다. 상관분석 결과를 보면 우선 광공업비율과 실업률이 유의한 상관관계가 있는 층은 강원읍면부, 충남읍면부, 경북읍면부, 경남읍면부 등 주로 읍면부 지역에 해당하는 4개 지역에서만 의미가 있었다. 광공업비율과 경활비율이 유의한 상관관계가 있는 층은 서울, 대구, 인천, 광주, 경기동부, 경기읍면부, 경남동부 등인 것으로 나

타났다. 광공업비율이 실업률과 상관관계가 높은 지역은 주로 읍면 지역인 것과 달리 광공업비율과 경황비율의 상관관계가 높은 지역은 도시 지역과 동부 지역이 포함 된다는 차이가 있다.

〈표 3-24〉 정렬기준변수와 실업률, 경황비율과의 상관계수

층	정렬변수	실업률	경황비율	층	정렬변수	실업률	경황비율
서울	광공업비율	0.0059	0.2306	충남동부	광공업비율	0.0532	0.2349
	농림어업비율	-0.1075	0.0284		농림어업비율	-0.0376	0.2476
	전월세가구비율	0.2953	0.2505		전월세가구비율	0.0480	0.0590
부산	광공업비율	-0.0136	0.0327	충남읍면부	광공업비율	0.4263	-0.0177
	농림어업비율	-0.1126	-0.1882		농림어업비율	-0.3082	0.3288
	전월세가구비율	0.1167	0.1828		전월세가구비율	0.1148	0.1181
대구	광공업비율	-0.0848	0.2009	전북동부	광공업비율	-0.1340	0.2736
	농림어업비율	-0.0885	0.1326		농림어업비율	0.0391	-0.0027
	전월세가구비율	0.3300	-0.0287		전월세가구비율	-0.1456	0.0250
인천	광공업비율	0.0484	0.2048	전북읍면부	광공업비율	0.0249	-0.1249
	농림어업비율	-0.2444	-0.1633		농림어업비율	0.0733	-0.0132
	전월세가구비율	0.1778	0.1615		전월세가구비율	0.1304	0.2311
광주	광공업비율	-0.0525	0.3092	전남동부	광공업비율	-0.1185	0.2669
	농림어업비율	0.0579	0.0994		농림어업비율	0.0023	0.1638
	전월세가구비율	0.3087	-0.2178		전월세가구비율	-0.0665	0.2223
대전	광공업비율	-0.0933	0.0050	전남읍면부	광공업비율	0.2750	-0.1487
	농림어업비율	0.0147	-0.0708		농림어업비율	-0.4629	0.1597
	전월세가구비율	0.2857	0.0383		전월세가구비율	0.3823	-0.0805
울산	광공업비율	-0.161	0.0046	경북동부	광공업비율	0.0213	0.2453
	농림어업비율	0.0175	0.0734		농림어업비율	-0.0349	0.2360
	전월세가구비율	0.1906	-0.0326		전월세가구비율	0.2319	0.0752
경기동부	광공업비율	0.0344	0.2274	경북읍면부	광공업비율	0.4191	-0.0978
	농림어업비율	-0.0944	0.0437		농림어업비율	-0.4319	0.2102
	전월세가구비율	0.0343	0.2557		전월세가구비율	0.3485	-0.0799
경기읍면부	광공업비율	0.0571	0.3371	경남동부	광공업비율	0.0099	0.4417
	농림어업비율	0.0343	-0.1032		농림어업비율	0.1064	-0.3219
	전월세가구비율	-0.0876	0.1523		전월세가구비율	0.2275	0.0361
강원동부	광공업비율	-0.1401	-0.0583	경남읍면부	광공업비율	0.3251	-0.0012
	농림어업비율	-0.0101	-0.306		농림어업비율	-0.3528	0.0746
	전월세가구비율	-0.094	0.2556		전월세가구비율	0.2818	-0.0117
강원읍면부	광공업비율	0.4075	0.0074	제주동부	광공업비율	0.1022	0.1001
	농림어업비율	-0.0091	-0.0626		농림어업비율	-0.1629	-0.2675
	전월세가구비율	-0.0914	-0.1526		전월세가구비율	0.4590	0.2896
충북동부	광공업비율	0.1473	0.2614	제주읍면부	광공업비율	0.0948	-0.0770
	농림어업비율	0.1456	0.1974		농림어업비율	-0.2827	0.4119
	전월세가구비율	0.1118	-0.0861		전월세가구비율	0.2544	-0.0107
충북읍면부	광공업비율	0.1566	0.2377				
	농림어업비율	-0.2611	-0.1257				
	전월세가구비율	0.1766	0.1563				

두 번째로 농림어업비율과 실업률이 유의한 상관관계가 있는 층은 인천, 충남읍면부, 전남읍면부, 경북읍면부, 경남읍면부 등으로 모두 음의 상관관계를 나타내고 있으며, 이는 농림어업비율이 높을수록 실업률은 떨어지는 것을 의미한다. 농림어업비율과 경황비율이 유의한 상관관계가 있는 층은 부산, 강원동부, 충남읍면부, 경남동부 등 네 곳으로 충남읍면부를 제외한 세 곳은 음의 상관관계를 나타내어 농림어업

비율이 높을수록 경찰비율이 낮아지는 것으로 나타났다.

세 번째로 전월세가구비율과 실업률의 상관관계가 유의하게 나타난 층은 서울 이외에도 대구, 대전, 전남읍면부, 경북읍면부, 제주동부 등이었고, 모두 양의 상관계수를 가지며 전월세가구비율이 높을수록 실업률이 높은 것으로 나타났다. 전월세가구비율과 경찰비율이 유의한 상관관계가 있는 층은 서울, 부산, 인천, 광주, 경기동부 등으로 광주를 제외한 나머지 지역은 모두 양의 상관관계를 가지며, 광주는 음의 상관관계를 갖는 것으로 나타났다.

이들 세 변수들과 실업률과 경찰비율에 대한 상관분석결과를 보면 광공업비율이나 농림어업비율은 생각보다 실업률이나 경찰비율과 통계적으로 유의적인 상관관계를 보이는 지역이 많지 않은 것으로 분석되었으며, 상대적으로 전월세비율은 더 많은 지역에서 실업률이나 경찰비율을 설명하는 변수로 의미가 있는 것으로 보인다.

<표 3-25> 거처유형과 실업률과 경찰비율과의 분산분석

층	실업률		경찰비율	
	F값	P값	F값	P값
서울	6.22	0.0004	13.58	<0.0001
부산	2.26	0.0840	2.11	0.1020
대구	1.10	0.3509	0.36	0.7797
인천	2.39	0.0721	4.07	0.0086
광주	0.25	0.8631	2.34	0.0781
대전	5.17	0.0023	1.03	0.3840
울산	2.2	0.0938	1.89	0.1365
경기동부	6.07	0.0006	16.47	<0.0001
경기읍면부	0.73	0.5392	0.94	0.4268
강원동부	1.08	0.3642	4.21	0.0095
강원읍면부	1.11	0.3513	1.32	0.2790
충북동부	0.25	0.8644	1.52	0.2213
충북읍면부	2.39	0.0785	0.52	0.6671
충남동부	3.45	0.0234	3.84	0.0150
충남읍면부	2.88	0.0443	0.42	0.7387
전북동부	1.67	0.1856	0.09	0.9659
전북읍면부	0.29	0.8309	0.6	0.6184
전남동부	2.03	0.1207	4.96	0.0042
전남읍면부	1.46	0.2354	0.12	0.9476
경북동부	1.25	0.2979	0.11	0.9550
경북읍면부	3.70	0.0159	0.87	0.4613
경남동부	3.02	0.0358	12.56	<0.0001
경남읍면부	5.67	0.0019	0.33	0.8061
제주동부	1.43	0.2517	0.89	0.4591
제주읍면부	0.03	0.8619	0.42	0.5250

한편 분류기준 변수 중 거처유형은 범주형 변수이기 때문에 실업률, 경찰비율과 분산분석을 통해 연관성을 검토하였으며, 분석결과는 <표 3-25>에 정리되어 있다. 거처유형과 실업률과의 분산분석 결과 서울, 대전, 경기동부, 충남동부, 충남읍면부,

경북읍면부, 경남동부, 경남읍면부에서 유의한 차이를 나타냈다. 아울러 거처유형과 경찰비율과의 분산분석 결과 서울, 인천, 경기동부, 강원동부, 충남동부, 전남동부, 전남읍면부, 경남동부에서 유의한 결과를 얻었다. 분석결과를 보면 상당히 많은 지역(층)에서 거처유형이 경제활동인구조사에서 핵심적인 변수인 실업률이나 경찰비율과 상관관계가 높은 것으로 나타났으며, 결과적으로 거처유형은 조사구 정렬기준 변수를 상당히 중요한 역할을 할 수 있다는 결론을 내릴 수 있다.

(2) 가계동향조사 주요 변수와 정렬기준과의 상관분석

가구부문 표본설계에서 사용하는 정렬기준 변수와 가계동향조사의 주요 변수인 경상소득, 근로소득, 사업소득, 교육, 주거 및 수도광열, 보건과 공공업비율, 농림어업비율, 전월세비율과 상관분석결과는 <표 3-26>과 같다.

분석결과를 보면 공공업비율과 가계자료 변수 간 상관관계가 유의한 지역은 대전, 울산, 경기동부, 경기읍면부, 강원읍면부, 충북읍면부, 경북동부, 경북읍면부, 경남읍면부, 제주동부 등이었고, 농림어업비율과 가계자료 변수 간 상관관계가 유의한 지역은 대구, 광주, 경기동부, 강원동부, 강원읍면부, 충북동부, 전북읍면부, 경북읍면부, 경남동부, 경남읍면부 등이었다. 전월세비율과 가계자료 변수 간 상관관계가 유의한 지역은 서울, 부산, 대구, 인천, 광주, 대전, 울산, 경기동부, 경기읍면부, 충북동부, 충남동부, 전남동부, 경북동부, 경북읍면부, 경남동부, 제주읍면부 등이었다.

공공업비율과 농림어업비율의 경우 경제활동인구조사 측면에서는 큰 역할을 한다고 볼 없지만 상대적으로 가계동향조사의 표본설계 측면에서 보면 정렬기준 변수로 활용될 수 있는 변수라고 볼 수 있다. 하지만 제시된 분석결과는 각 변수를 독립적으로 분석한 결과로 공공업비율이나 농림어업비율이 서로 상관관계를 갖고 있을 수 있기 때문에 이들 변수와 다양한 가계동향조사 변수를 통계적인 모형을 이용해 종합적으로 분석하는 것이 필요하다. 이와 관련된 분석은 2절에서 심층적으로 살펴 보기로 한다.

〈표 3-26〉 정렬기준변수와 가계자료 주요변수와의 상관계수

		경상소득	근로소득	사업소득	교육	주거 및 수도광열	보건
서울	광공업비용	-0.0475	-0.0724	0.0140	-0.1375	-0.0302	-0.0541
	농림어업비용	-0.0796	-0.0471	-0.0677	-0.0866	-0.0361	-0.0891
	전월세가구비용	-0.4069	-0.3556	-0.1966	-0.3168	-0.2079	-0.2385
부산	광공업비용	-0.0747	-0.0948	0.1098	0.0133	-0.1890	-0.1757
	농림어업비용	0.0393	-0.0758	0.0926	-0.0522	-0.1951	0.0239
	전월세가구비용	-0.2485	-0.1347	-0.1910	-0.3102	-0.0387	-0.2196
대구	광공업비용	0.0753	0.1048	0.0047	0.1294	-0.2180	-0.0928
	농림어업비용	-0.2618	-0.2170	-0.1236	-0.2119	-0.1843	-0.0858
	전월세가구비용	-0.3303	-0.2012	-0.2247	-0.3665	0.0022	-0.0184
인천	광공업비용	-0.0988	-0.0214	-0.0774	-0.2280	-0.3548	0.1128
	농림어업비용	-0.0584	-0.0785	0.0412	-0.1050	0.1673	0.0049
	전월세가구비용	-0.5849	-0.5623	-0.0647	-0.6165	-0.0569	-0.0791
광주	광공업비용	-0.1530	-0.0339	-0.1270	0.0224	-0.1105	-0.1679
	농림어업비용	-0.1404	-0.0802	-0.2076	-0.0511	-0.2667	-0.2075
	전월세가구비용	-0.3997	-0.2676	-0.3480	-0.3624	-0.0155	-0.2837
대전	광공업비용	-0.1690	-0.1170	-0.1068	-0.2245	-0.2883	-0.0174
	농림어업비용	-0.1489	-0.2357	0.0582	-0.2082	0.1327	0.1670
	전월세가구비용	-0.3750	-0.3301	-0.1276	-0.2575	-0.1509	-0.2001
울산	광공업비용	0.3822	0.5396	-0.3513	0.1207	-0.1727	0.3264
	농림어업비용	-0.013	-0.0085	0.0069	0.1429	-0.2155	-0.1059
	전월세가구비용	-0.5552	-0.5175	-0.0278	-0.4602	0.2709	-0.1774
경기동부	광공업비용	-0.2607	-0.2845	0.0577	-0.1884	-0.2043	-0.1916
	농림어업비용	-0.2143	-0.1961	-0.0672	-0.2861	-0.0240	0.1689
	전월세가구비용	-0.4656	-0.3814	-0.2341	-0.3975	-0.1894	-0.0143
경기읍면부	광공업비용	0.2712	0.3819	-0.1447	0.3067	-0.0403	-0.0209
	농림어업비용	-0.0909	-0.2012	-0.0102	-0.1069	-0.0881	-0.0244
	전월세가구비용	-0.1552	-0.1902	0.1756	-0.4246	0.0773	-0.0007
강원동부	광공업비용	0.1145	-0.0618	0.1435	0.0019	-0.0245	0.0729
	농림어업비용	-0.4351	-0.3622	-0.1337	-0.3524	-0.3184	0.0473
	전월세가구비용	0.2784	0.2462	0.2301	-0.0633	0.0369	-0.0636
강원읍면부	광공업비용	0.1464	0.1737	-0.1147	0.1352	-0.1990	0.5339
	농림어업비용	-0.3817	-0.3159	-0.0670	-0.1562	-0.2014	-0.2196
	전월세가구비용	0.1458	0.1824	0.0040	-0.1425	0.0066	0.1384
충북동부	광공업비용	0.0390	0.1552	-0.1022	0.1064	-0.1524	0.0042
	농림어업비용	-0.2732	-0.3913	0.1067	-0.2740	0.2707	0.0489
	전월세가구비용	-0.3671	-0.4667	0.2404	-0.2278	0.0815	-0.3430
충북읍면부	광공업비용	0.2653	0.1449	0.1814	0.4202	0.0393	0.2400
	농림어업비용	-0.179	-0.0231	-0.3464	-0.3612	-0.1005	-0.2006
	전월세가구비용	0.0674	0.0499	0.1171	0.0945	0.2193	0.1343
충남동부	광공업비용	0.0096	0.2520	-0.2274	0.1080	0.0337	0.1344
	농림어업비용	-0.2543	-0.1966	-0.0453	-0.1781	-0.1621	-0.0903
	전월세가구비용	-0.3862	-0.2710	-0.1882	-0.3498	-0.1760	-0.3100
충남읍면부	광공업비용	-0.0057	0.0593	-0.1122	0.1402	0.0615	-0.0227
	농림어업비용	-0.3132	-0.2506	-0.1199	-0.2251	-0.2899	-0.1542
	전월세가구비용	0.0733	0.1311	-0.1322	0.3335	0.0954	-0.0617
전북동부	광공업비용	-0.0341	0.3090	-0.2627	-0.0234	-0.1559	-0.1524
	농림어업비용	-0.0338	-0.1159	0.1533	-0.0056	-0.1509	-0.0794
	전월세가구비용	-0.1843	-0.2931	0.0829	-0.0401	-0.0131	-0.1557
전북읍면부	광공업비용	0.1911	0.2759	0.1566	0.3537	0.0770	-0.1531
	농림어업비용	-0.6189	-0.6604	-0.3955	-0.5564	-0.4418	0.2678
	전월세가구비용	0.3352	0.3958	0.2124	0.1842	0.2793	-0.1936
전남동부	광공업비용	0.2350	0.2233	0.0619	-0.0651	-0.0901	0.0711
	농림어업비용	-0.2110	-0.2151	-0.0574	-0.1342	-0.1939	-0.1263
	전월세가구비용	0.1009	0.0680	0.1813	-0.0941	-0.0793	0.3069

		경상소득	근로소득	사업소득	교육	주거 및 수도광열	보건
전남읍면부	광공업비율	0.1043	-0.0636	0.2432	0.1643	0.0571	-0.2856
	농림어업비율	-0.3095	-0.3150	-0.0865	-0.4417	-0.0987	0.0061
	전월세가구비율	-0.0247	-0.0125	0.0527	-0.0452	0.2235	-0.2277
경북동부	광공업비율	0.4283	0.5159	-0.0996	0.4429	0.0415	-0.0412
	농림어업비율	-0.3125	-0.2685	-0.0377	-0.1920	-0.0725	-0.0812
	전월세가구비율	-0.3451	-0.3274	0.0191	-0.2203	0.1086	-0.1088
경북읍면부	광공업비율	0.3506	0.4497	-0.0172	0.4290	0.4329	0.1262
	농림어업비율	-0.7179	-0.6965	-0.3458	-0.6529	-0.5266	-0.1732
	전월세가구비율	0.3076	0.3648	-0.0036	0.2269	0.3159	-0.1804
경남동부	광공업비율	0.1594	0.2939	-0.0496	0.0256	-0.0028	-0.1666
	농림어업비율	-0.3275	-0.3633	-0.0412	-0.2026	-0.1542	0.0303
	전월세가구비율	-0.4334	-0.2795	-0.3281	-0.3610	0.1350	-0.3182
경남읍면부	광공업비율	0.3679	0.6183	-0.3419	0.2711	0.0656	-0.1423
	농림어업비율	-0.4192	-0.3957	0.0646	-0.3297	0.0718	-0.2565
	전월세가구비율	-0.0232	-0.2060	0.3138	0.0691	-0.0968	0.0738
제주동부	광공업비율	0.4178	0.4104	0.0648	0.0648	0.5045	0.2976
	농림어업비율	-0.3205	-0.2841	-0.4036	-0.1221	-0.2980	-0.2202
	전월세가구비율	-0.0007	0.0930	-0.1110	-0.3432	0.1226	-0.1839
제주읍면부	광공업비율	-0.2885	-0.4211	-0.0030	0.2997	0.0027	-0.1203
	농림어업비율	-0.0888	-0.2923	-0.1160	-0.5465	0.2850	-0.0430
	전월세가구비율	0.6399	0.3180	0.6001	0.7484	0.3269	-0.0751

한편, 거처유형에 따른 경상소득, 근로소득, 사업소득, 교육, 주거 및 수도광열, 보건비 등에 대한 분산분석 결과, 거처유형에 따른 경상소득과의 분산분석 결과는 대부분의 지역에서 유의한 것으로 나타났다.

거처유형과 사업소득과의 분산분석 결과 서울, 부산, 대구, 인천, 전북읍면부, 경남동부에서 유의한 결과를 얻었다. 또한 거처유형과 교육비와의 분산분석결과는 대부분의 층에서 유의한 결과를 얻었고, 거처유형과 주거 및 수도광열비와의 분산분석 결과에서도 서울, 부산, 대전, 울산, 경기동부, 충남동부에서 유의한 결과를 얻었으며, 거처유형과 보건비와의 분산분석 결과 서울, 대구, 광주, 충북동부에서 유의한 결과를 얻었다. 결과적으로 거처유형은 경제활동인구조사에서와 마찬가지로 가계동향조사의 주요 변수와도 연관성이 매우 높은 것으로 판단되면 따라서 가계동향조사 관점에서도 정렬기준 변수로 중요한 역할을 하고 있다고 보인다.

〈표 3-27〉 거처유형과 가계자료 주요변수와의 분산분석

	경상소득		근로소득		사업소득		교육비		주거및수도광열비		보건비	
총	F값	P값	F값	P값	F값	P값	F값	P값	F값	P값	F값	P값
서울	51.36	<.0001	38.06	<.0001	3.76	0.0122	27.23	<.0001	24.40	<.0001	8.87	<.0001
부산	6.18	0.0007	3.05	0.0326	5.84	0.0011	4.66	0.0045	5.03	0.0029	1.46	0.2298
대구	9.24	<.0001	3.90	0.0127	4.49	0.0064	7.36	0.0003	1.13	0.3417	2.82	0.0459
인천	9.62	<.0001	5.15	0.0030	3.26	0.0275	8.14	0.0001	1.52	0.2178	0.70	0.5562
광주	13.95	<.0001	5.38	0.0023	2.49	0.0684	6.19	0.0009	1.63	0.1902	5.62	0.0017
대전	23.99	<.0001	17.21	<.0001	1.65	0.1855	7.78	0.0002	5.94	0.0012	1.73	0.1692
울산	4.70	0.0065	4.49	0.0080	1.62	0.1992	2.83	0.0500	4.42	0.0086	0.19	0.9007
경기/ 동부	26.78	<.0001	20.84	<.0001	1.65	0.1835	14.08	<.0001	10.12	<.0001	1.98	0.1222
경기/ 읍면	3.30	0.0300	2.26	0.0962	1.01	0.4001	1.66	0.1913	0.83	0.4878	0.29	0.8323
강원/ 동부	3.93	0.0146	4.42	0.0085	0.31	0.8145	3.55	0.0221	0.86	0.4689	0.10	0.9595
강원/ 읍면	0.95	0.4296	1.74	0.1794	0.64	0.5927	0.87	0.4652	0.62	0.6085	1.58	0.2145
충북/ 동부	5.34	0.0047	4.35	0.0120	0.05	0.9853	3.15	0.0400	2.08	0.1240	3.22	0.0370
충북/ 읍면	1.60	0.2100	0.68	0.5713	1.14	0.3495	2.07	0.1251	0.40	0.7554	1.83	0.1614
충남/ 동부	7.70	0.0004	3.23	0.0327	2.76	0.0547	3.20	0.0337	3.27	0.0314	1.07	0.3716
충남/ 읍면	0.41	0.7456	0.20	0.8923	1.00	0.4069	1.87	0.1566	0.70	0.5596	0.64	0.5949
전북/ 동부	4.11	0.0122	6.83	0.0008	0.99	0.4089	2.73	0.0562	2.35	0.0860	0.08	0.9683
전북/ 읍면	5.78	0.0052	4.02	0.0218	4.79	0.0113	21.98	<.0001	2.59	0.0816	0.42	0.7375
전남/ 동부	8.11	0.0010	6.17	0.0045	1.58	0.2190	3.12	0.0546	2.62	0.0850	1.95	0.1553
전남/ 읍면	0.33	0.8008	0.36	0.7795	0.52	0.6717	3.44	0.0363	1.11	0.3694	1.40	0.2707
경북/ 동부	5.49	0.0029	3.26	0.0310	0.24	0.8689	6.05	0.0017	0.02	0.9949	2.83	0.0501
경북/ 읍면	9.57	0.0001	16.20	<.0001	0.34	0.7958	10.42	<.0001	2.19	0.1081	0.73	0.5418
경남/ 동부	5.01	0.0047	2.65	0.0614	4.99	0.0047	2.25	0.0962	0.37	0.7754	1.22	0.3135
경남/ 읍면	2.09	0.1212	4.68	0.0082	0.97	0.4177	1.93	0.1449	0.25	0.8592	0.28	0.8378
제주/ 동부	1.43	0.2649	1.69	0.2010	0.35	0.7867	3.90	0.0242	0.35	0.7903	0.26	0.8531
제주/ 읍면	0.10	0.7628	0.26	0.6267	0.15	0.7114	8.24	0.0208	0.23	0.6414	2.56	0.1485

4) 업무량 파악을 위한 FGI 결과

현행 연동 모형을 개편하는 경우 최대 쟁점 중 하나는 늘어나는 업무량을 현재 통계청 지방 실사 담당 인력으로 처리할 수 있는지, 만약 현재의 인력으로 부족하다면 어느 정도 충원하는 것이 필요한지를 파악하는 것이다. 아울러 향후 업무량 조정 및 인력 확충을 위해서는 현행 경제활동인구조사 및 가계동향조사에서 조사 과정별로 어떤 업무에 어느 정도의 시간이 소요되고, 어떤 업무처리에 조사직원들이 가장 어려움을 느끼는지 등을 구체적으로 파악하는 것이 필요하다.

이런 관점에서 본 연구에서는 가구부문 경성조사 연동모형 개선 관련 FGI(표적집 단면접)를 수행하였다. FGI는 현장 조사 담당자와의 자유로운 토론을 통해 가구부문 조사현장의 실태 및 현행 연동모형의 한계점과 개선 방안을 도출하기 위한 정보 수집을 목적으로 이루어졌다. 여러 가지 사정상 본 FGI는 연구진이 직접 진행(moderating)하는 대신 2011년 9월 30일(금) 10시30분부터 14시30분까지 지방청 및 사무소 직원 11명을 대상으로 본 연구진이 개발한 FGI 진행 설문지를 이용해 통계청 표본과 주관으로 진행되었다. 본 연구에서 지방 실사 담당자를 대상으로 수행된 FGI를 통해 수렴된 의견을 정리한 결과는 다음과 같다.

(1) 업무량 수준

가) 경제활동인구조사 및 가계동향 조사 업무량

⇒ 대부분 경찰 3개/가계 2개 조사구 담당

○ 경찰 및 가계 업무 이외에 담당 업무

: 무기직 경우는 이외 업무가 없으나, 정규직 공무원 대부분의 경우는 있음

⇒ 행정, 가구관련 조사, 타 조사업무 관련 실사지도 및 총괄업무 등

○ 현재 담당하고 있는 다른 업무를 모두 줄인다면, 어느 정도 경찰 및 가계동향 조사 업무를 늘리는 것이 가능한지 여부

⇒ 대부분 불가능 (1명만 경찰 1개 조사구 정도 추가 가능하다고 응답함)

나) 경찰 40가구(2개 조사구)와 가계동향 10가구(1개 조사구) 업무량 비교

⇒ 경찰 40가구(2개 조사구) > 가계 10가구(1개 조사구) > 경찰 20가구 순

· 경찰의 업무량 과다는 부가조사(3월*, 5월, 6월*, 8월, 9월*, 12월*) 등에 기인함(*는 지역별 고용조사 분기 실시월임)

(2) 조사 과정별 업무 난이도

가) 경찰이나 가계조사 과정 중 가장 어려운 업무

⇒ 가구설득, 부재가구 재접촉, 유대관계 형성, 가구명부 작성, 부가조사 순

나) 경찰이나 가계조사 과정 중 시간이 가장 많이 소요되는 업무

⇒ 가계 및 경찰조사 공통 : 불응가구 설득 친밀감 형성

⇒ 가계조사 : 가계부 기입 설득 및 작성지도, 누락항목 보완

⇒ 경찰조사 : 가구원별 조사 및 부가조사

(3) 신규 (연동) 조사구의 가구 조사 참여 설득 과정

가) 조사구 연동에 따라 새로운 조사구로 조사대상이 변경되는 경우 새로운 조사구 명부 작성과 조사구내 표본 가구(20가구)를 대상으로 경찰 조사 참여를 설득하기 위해 최소한 몇 번 정도 조사구 현지 방문이 필요합니까?

○ 현지방문 필요 횟수

⇒ 신규 조사구 조사 참여 설득을 위해 평균 21.5회 방문

· 60가구 명부작성 5.5일+20가구 가구명부 10일+예비조사 3일+본조사 3일

○ 최소한 얼마 동안의 기간이 필요하다고 생각하십니까?

⇒ 필요 기간 : 평균 6.7개월 소요

나) 그동안 경험에 의하면 조사 참여 설득을 위해 보통 몇 번 정도 가구를 방문하는 것이 필요한지요? 최대 몇 번까지 가구를 방문해 본 경험이 있습니까? 아울러 어느 정도 인터뷰를 두고 방문해 설득하는 것이 효과적이라고 생각하십니까?

○ 평균 가구 방문 횟수 : 평균 5.4회 (4번 3명, 5번 3명 등)

○ 최대 가구 방문 횟수 : 평균 8.9회 (8번 2명, 10번 2명 등)

○ 가구 불응시 어느 정도 기간을 두고 방문하여야 효과적인가

⇒ 평균 약 3주 (1주일 : 3명, 1달 : 3명 등)

다) 연동에 따른 조사구 교체 업무 없이 고정된 조사구를 대상으로 조사를 한다면 현행 담당하는 조사 업무량(경찰 60가구, 가계동향 20가구)에 어느 정도 추가 업무를 수행하는 것이 가능합니까?

⇒ 대부분 불가능 (경찰 부가 조사에 따른 조사업무 증가, 조사업무의 입력·내검 등의 물리적 시간 부족, 불응가구 증가 때문)

(4) 친밀감(rapport) 형성에 걸리는 시간

가) 처음 조사에 참여한 가구의 경우, 보통 어느 정도 기간 동안에 걸쳐 몇 번 정도의 방문이 이루어지면 가구 방문시 커피나 다과를 대접 받을 수 있을 정도로 친밀한 관계(rapport) 형성이 가능하다고 생각하십니까?

⇒ 기간 : 약 5.6개월(3~6개월 5명, 6개월이상 4명 등)

⇒ 횟수 : 5~10회미만 2명, 10~15회미만 4명 등

나) 경찰조사에서 조사 참여 초기에 가구원 중 실직자가 있는데(생겼는데) 의도적으로 이런 사실을 숨기고 응답을 제대로 하지 않는 경우 어느 정도의 시간이 경과해 서로 친숙해지면 이런 숨김없이 솔직한 답변을 얻을 수 있었습니까?

⇒ 평균 약 6.8개월(3~6개월미만 3명, 6개월이상 3명 등)

다) 가계동향조사에서 조사 참여 초기에 일부 지출 항목이나 소득 항목을 응답자가 의도적으로 숨기고 없다고 응답하는 경우 어느 정도의 시간이 경과해 서로 친숙해지면 이런 숨김없이 솔직한 답변을 얻을 수 있었습니까?

⇒ 평균 약 10.5개월(3~6개월미만 2명, 6개월 2명, 1년 3명 등)

(5) 매월 장기조사에 따른 응답 피로 현상

가) 어떤 가구가 응답하기 귀찮아서 가계부를 불성실하게 작성한 사례가 있다면 평균적으로 몇 개월 정도 조사에 참여한 후 이런 현상이 나타났습니까?

⇒ 평균 약 18.4개월(1년 미만 2명, 1~2년 5명 등)

나) 만약 지난 달과 거의 동일한 응답을 습관적으로 한다고 느끼는 경우, 보다 성실한 응답을 얻기 위해 개인적으로 어떻게 대응을 하십니까?

⇒ 모르는 척 일부 상황에 대해 다르게 문의

⇒ 사적인(일상적인) 얘기를 하면서 도중에 물어봄

⇒ 전화 또는 문자메시지 이용 등

(6) 연동모형 변경에 따른 업무량 파악 관련

가) 현행 조사방식에서 조사구 교체(연동)가 이루어지는 기간 동안 전체 경찰 및 가계동향조사 업무량 중 조사구 교체로 인해 발생하는 업무량은 어느 정도

비율을 차지합니까? 36.4%(30% 6명, 40% 3명 등)

※ 경찰 60, 가계동향 20가구 조사하는 보통의 업무량 기준임, 이외의 경상(기타) 업무는 제외

나) 조사구별로 현행 경찰 및 가계동향조사 업무(경찰 20가구 가계동향 10가구) 대신 **경찰 조사의 경우 조사구당 15가구** 가계동향조사는 종전대로 10가구를 조사한다면, 현재 업무량에 비해 어느 정도 **업무량이 줄어들 것** 같습니다?
⇒ 평균 5.9% 감소(5%미만 2명, 10~15% 5명 등)

다) 현행 같이 조사구당 4개 구역에서 20가구(가계동향은 10가구)를 조사하는 방식 대신 위와 같이 **3개 구역 15가구(가계동향은 10가구)를 조사한다면**, 조사 업무 수행상 어떤 **문제나 불편함이 생길 수 있는지**?
⇒ 조사구 수가 늘어나지 않으면 괜찮으나, 조사구 수가 증가하면 업무 부담이 높아짐(조사구 관리 및 이동 시간), 아파트의 경우 옆집이 조사 안할 경우 불응 가능성이 높아짐 등

2. 연동모형 개선안

1) 총화 및 정렬

(1) 추가 정렬기준변수의 검토

향후 가구부문 표본설계 개선방향을 모색하기 위해, 현재 경제활동인구조사 등 통계청 가구부문 표본설계에서 사용하고 있는 정렬기준 변수 외에 일부 변수를 정렬기준 변수를 활용하는 방안을 검토하고자 한다. 연구진들의 그동안 경찰 관련 자료 분석 경험에 의하면 조사구별 30~40대 비율과 30대 이상 대졸자비율이 조사구별 실업률과 상관관계가 높은 것으로 추측되어 이들 2개 변수와 경찰조사의 핵심 변수인 실업률, 경찰비율과의 상관분석을 추가적으로 실시하였다.

2005년 인구주택총조사 시점에 조사구별 30~40대 비율과 실업률, 경찰비율에 대한 상관분석 결과를 정리해 보면 <표 3-28>과 같다. 서울, 대구, 대전, 울산, 강원읍면부, 전북동부, 전남동부, 경북읍면부, 경남읍면부 등에서 30~40대 비율과 실업률이 유의한 음의 상관관계가 있었고, 서울, 광주, 경기동부, 강원동부, 충북읍면부에서

30~40대 비율과 경찰비율간 유의한 양의 상관관계가 있다는 것을 확인할 수 있다. 여기서 30~40대 비율이 높은 조사구는 상대적으로 경제활동에 참여할 의사가 높은 청년층이 다수 포함된 조사구를 의미하며, 상관분석 결과에서 볼 수 있는 것처럼 청년층의 비율과 해당 조사구의 실업률과 경찰비율과 상당한 연관성이 있음을 확인할 수 있다.

한편 25개 층별로 30대 이상 대졸자비율과 2010년도 경제활동인구 조사에서 실업률 및 경찰비율과의 상관분석 결과는 <표3-29>와 같다. 상관분석 결과 상당히 많은 지역에서 30대 이상 대졸자비율과 실업률 및 경찰비율과 유의적인 상관관계를 갖는 것으로 분석되었다.

<표 3-28> 30~40대 비율과 실업률, 경찰비율과의 상관계수

	실업률	경찰비율
서울	-0.1895	0.1407
부산	-0.0104	0.1393
대구	-0.2551	0.1632
인천	-0.0847	0.1491
광주	-0.1812	0.3317
대전	-0.3036	0.1897
울산	-0.3715	0.1464
경기동부	-0.0927	0.1551
경기읍면부	-0.0213	0.1484
강원동부	-0.0441	0.4746
강원읍면부	-0.2817	0.0448
충북동부	-0.0438	0.2072
충북읍면부	0.1656	0.2684
충남동부	-0.2129	0.0944
충남읍면부	0.1738	0.0090
전북동부	-0.2979	0.0929
전북읍면부	0.0407	0.1069
전남동부	-0.3438	0.3919
전남읍면부	0.1711	0.0414
경북동부	-0.1597	0.1998
경북읍면부	-0.1895	0.1407
경남동부	-0.0104	0.1393
경남읍면부	-0.2551	0.1632
제주동부	-0.0847	0.1491
제주읍면부	-0.0343	0.1124

〈표 3-29〉 30대 이상 대졸자비율과 실업률, 경찰비율과의 상관계수

층	실업률	경찰비율
서울	-0.3119	-0.3320
부산	-0.0604	-0.1768
대구	-0.2293	0.0315
인천	-0.2672	0.0027
광주	-0.3820	0.1241
대전	-0.3517	-0.0146
울산	-0.2779	-0.0525
경기동부	-0.2397	-0.3770
경기읍면부	-0.0966	0.0444
강원동부	-0.0673	0.3506
강원읍면부	-0.3400	-0.0744
충북동부	-0.1051	-0.1939
충북읍면부	0.0415	0.4010
충남동부	-0.2938	-0.1131
충남읍면부	0.2805	-0.1677
전북동부	-0.2157	-0.1521
전북읍면부	-0.0644	0.1038
전남동부	-0.1761	0.2009
전남읍면부	0.1649	-0.0032
경북동부	-0.2967	-0.0401
경북읍면부	0.3643	-0.0083
경남동부	-0.3173	0.0469
경남읍면부	0.1305	-0.0623
제주동부	0.0916	0.3772
제주읍면부	0.3743	0.1328

지금까지 제시된 기존 정렬변수와 추가로 고려한 2개 변수와 실업률 및 경찰비율과의 상관관계 분석결과를 종합적으로 살펴보기로 한다. 이를 위해 전체 조사구를 서울, 광역시, 동부, 읍면부 등 4개 범주로 구분한 후 기존 정렬변수, 30~40대비율, 30대 이상 대졸자비율과 경찰 핵심변수인 실업률, 경찰비율 및 가계동향조사 주요 변수들과의 상관분석을 수행하였다. 상관분석결과를 정리한 결과는 <표 3-30> 및 <표 3-31>과 같다.

〈표 3-30〉 4개 지역에서 실업률, 경찰비율과 거주유형을 제외한 정렬기준변수와의 상관계수

			광공업 비율	농림어업 비율	전월세 비율	30~40대 비율	30대이상 대졸자비율
서울	실업률	상관계수	0.0059	-0.1075	0.2953	-0.1895	-0.3119
		p값	0.9283	0.1023	<.0001	0.0024	<.0001
		N	232	232	254	254	254
	경찰비율	상관계수	0.2306	0.0284	0.2505	0.1407	-0.3320
		p값	0.0004	0.6672	<.0001	0.0249	<.0001
		N	232	232	254	254	254
광역시	실업률	상관계수	-0.0125	-0.0793	0.2245	-0.1493	-0.2530
		p값	0.7620	0.0549	<.0001	0.0001	<.0001
		N	586	586	655	655	655
	경찰비율	상관계수	0.1747	0.0137	-0.0017	0.2274	-0.0453
		p값	<.0001	0.7407	0.9661	<.0001	0.2472
		N	586	586	655	655	655
동부	실업률	상관계수	0.0690	-0.0559	0.0685	-0.0525	-0.1532
		p값	0.1108	0.1966	0.0899	0.1937	0.0001
		N	536	536	615	615	615
	경찰비율	상관계수	0.2515	0.0425	0.1261	0.2778	-0.0654
		p값	<.0001	0.3262	0.0017	<.0001	0.1053
		N	536	536	615	615	615
읍면부	실업률	상관계수	0.3064	-0.3092	0.1816	0.2038	0.1234
		p값	<.0001	<.0001	<.0001	<.0001	0.0059
		N	437	437	497	497	497
	경찰비율	상관계수	-0.0189	0.1708	-0.0108	0.0147	0.0111
		p값	0.6934	0.0003	0.8095	0.7444	0.8050
		N	437	437	497	497	497

우선 <표 3-30>을 보면 기존 정렬 변수 중 전월세비율과 추가로 고려한 30~40대 비율 및 30대 이상 대졸자비율이 광공업비율이나 농림어업비율에 비해 상대적으로 상관관계가 높은 것으로 나타났다.

한편 <표 3-31>의 가계동향조사 자료의 주요 변수와 기존 정렬변수 및 추가로 고려한 2개 변수에 대한 상관관계 분석결과를 보면, 전반적으로 전월세비율과 30~40대 비율, 30대 이상 대졸자비율이 상당히 의미 있는 정렬기준 변수가 될 수 있다는 것을 알 수 있으며, 서울의 경우 광공업비율이나 농림어업비율은 별 의미가 없다는 것을 확인할 수 있다. 따라서 향후 가구부분 표본설계에 있어서는 기존 정렬변수들과 함께 본 연구에서 추가로 제안한 2개의 변수를 적극적으로 활용하는 방안을 고려할 필요가 있을 것으로 판단된다.

〈표 3-31〉 4개 지역에서 거쳐유형을 제외한 정렬기준변수와 가계자료 주요변수와의 상관계수

			경상소득	근로소득	사업소득	교육	주거및 수도광열	보건
서울	광공업 비율	상관계수	-0.0475	-0.0724	0.0140	-0.1375	-0.0302	-0.0541
		p값	0.5734	0.3904	0.8682	0.1015	0.7205	0.5212
		N	143	143	143	143	143	143
	농림어업 비율	상관계수	-0.0796	-0.0471	-0.0677	-0.0866	-0.0361	-0.0891
		p값	0.3445	0.5768	0.4215	0.3040	0.6686	0.2902
		N	143	143	143	143	143	143
	전월세 비율	상관계수	-0.4069	-0.3556	-0.1966	-0.3168	-0.2079	-0.2385
		p값	<.0001	<.0001	0.0142	<.0001	0.0095	0.0028
		N	155	155	155	155	155	155
	30~40대 비율	상관계수	0.0279	0.0498	0.0179	-0.0125	-0.0354	-0.0332
		p값	0.7309	0.5383	0.8254	0.8769	0.6617	0.6817
		N	155	155	155	155	155	155
	30대이상 대출자비율	상관계수	0.5802	0.5342	0.2460	0.5560	0.3959	0.2924
		p값	<.0001	<.0001	0.0020	<.0001	<.0001	0.0002
		N	155	155	155	155	155	155
광역시	광공업 비율	상관계수	0.0563	0.1817	-0.0885	0.0113	-0.1404	0.0207
		p값	0.2879	0.0006	0.0946	0.8320	0.0078	0.6959
		N	358	358	358	358	358	358
	농림어업 비율	상관계수	-0.1307	-0.1273	-0.0361	-0.1081	-0.0806	-0.0537
		p값	0.0133	0.0159	0.4959	0.0409	0.1280	0.3107
		N	358	358	358	358	358	358
	전월세 비율	상관계수	-0.3863	-0.3107	-0.1879	-0.3662	-0.0096	-0.1751
		p값	<.0001	<.0001	0.0002	<.0001	0.8482	0.0004
		N	401	401	401	401	401	401
	30~40대 비율	상관계수	0.2990	0.3607	0.0252	0.3111	0.0702	0.0927
		p값	<.0001	<.0001	0.6145	<.0001	0.1605	0.0638
		N	401	401	401	401	401	401
	30대이상 대출자비율	상관계수	0.5664	0.4541	0.2201	0.5193	0.2811	0.1744
		p값	<.0001	<.0001	<.0001	<.0001	<.0001	0.0005
		N	401	401	401	401	401	401
동부	광공업 비율	상관계수	0.1657	0.2179	0.0095	0.1286	-0.0295	0.0137
		p값	0.0016	<.0001	0.8577	0.0147	0.5776	0.7960
		N	359	359	359	359	359	359
	농림어업 비율	상관계수	-0.2443	-0.2408	-0.0363	-0.1704	-0.1107	-0.0756
		p값	<.0001	<.0001	0.4932	0.0012	0.0360	0.1530
		N	359	359	359	359	359	359
	전월세 비율	상관계수	-0.2315	-0.2006	-0.0546	-0.2375	-0.0126	-0.0967
		p값	<.0001	<.0001	0.2712	<.0001	0.8002	0.0510
		N	408	408	408	408	408	408
	30~40대 비율	상관계수	0.3974	0.3953	0.1237	0.3632	0.1063	0.0343
		p값	<.0001	<.0001	0.0124	<.0001	0.0318	0.4893
		N	408	408	408	408	408	408
	30대이상 대출자비율	상관계수	0.6061	0.5449	0.1419	0.5526	0.2666	0.1486
		p값	<.0001	<.0001	0.0041	<.0001	<.0001	0.0026
		N	408	408	408	408	408	408
읍면부	광공업	상관계수	0.2920	0.3837	-0.0811	0.3344	0.1296	0.0345

			경상소득	근로소득	사업소득	교육	주거및 수도광열	보건
	비율	p값	<.0001	<.0001	0.2165	<.0001	0.0477	0.5993
		N	234	234	234	234	234	234
	농림어업 비율	상관계수	-0.4322	-0.3960	-0.1483	-0.3981	-0.2139	-0.1587
		p값	<.0001	<.0001	0.0233	<.0001	0.0010	0.0151
		N	234	234	234	234	234	234
	전월세 비율	상관계수	0.1233	0.0972	0.1105	0.0649	0.1600	0.0147
		p값	0.0432	0.1117	0.0703	0.2892	0.0086	0.8107
		N	269	269	269	269	269	269
	30~40대 비율	상관계수	0.4835	0.4722	0.1123	0.4917	0.1560	-0.0105
		p값	<.0001	<.0001	0.0659	<.0001	0.0104	0.8638
		N	269	269	269	269	269	269
	30대이상 대졸자비율	상관계수	0.4295	0.3610	0.1569	0.4983	0.1760	0.0685
		p값	<.0001	<.0001	0.0100	<.0001	0.0038	0.2631
		N	269	269	269	269	269	269

(2) 정렬변수와 실업률과의 회귀분석

앞에서 언급한 것처럼 정렬기준 변수들은 서로 연관성을 갖고 있기 때문에 상관 분석을 통해서는 단편적인 변수들 간의 관계만을 확인하는 것이 가능하다. 따라서 기존 정렬변수들과 함께 추가로 제안한 30~40비율, 30대 이상 대졸자비율 등의 변수들 중 어떤 변수를 선택해 표본설계에 반영하는 것이 효과적인지 종합적으로 검토해 보기 위해, 여기서는 활용 가능한 정렬 기준 변수들을 설명변수로 사용하고, 실업률을 종속변수로 한 아래와 같은 다중선형회귀모형(multiple linear regression model)을 적용하여 분석해 보기로 한다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \epsilon$$

여기서 x_1 은 광공업비율, x_2 는 농림어업비율, x_3 는 전월세비율, x_4 는 30~40대 비율, x_5 는 30대 이상 대졸자비율 변수이며, 제시된 회귀분석모형에 대한 분석 결과는 <표 3-32>에 정리되어 있다.

참고로, 기존 정렬변수 중 거처유형은 범주형 변수로 거처유형을 포함한 일반화선형모형(GLM) 분석도 하였으나, 거처유형을 포함하지 않는 모형과 포함한 모형 모두에서 다른 변수에 대한 분석결과에 있어서 큰 차이가 없어서 여기서는 거처유형은 분석대상에서 제외한 모형을 사용하기로 한다.

서울은 광공업비율을 제외한 나머지 변수 모두 유의한 결과를 얻었고, 대구는 전월세비율, 인천은 농림어업비율과 30대 이상 대졸자비율, 광주 30대 이상 대졸자비율, 울산은 30~40대 비율, 경기동부 30대 이상 대졸자비율, 강원읍면부는 광공업비

율과 30~40대 비율, 충북동부는 30대 이상 대졸자비율, 충남읍면부는 광공업비율, 30~40대 비율, 30대 이상 대졸자비율이 유의한 변수로 분석되었다. 한편 전남동부는 30~40대 비율, 전남읍면부는 농림어업비율, 경남동부는 30대 이상 대졸자비율, 제주 동부는 30~40대 비율, 제주읍면부도 30~40대 비율이 유의한 변수인 것으로 나타났다.

<표 3-32>의 회귀분석 결과를 보면, 기존 정렬변수 외에 30~40대 비율과 30대 이상 대졸자비율이 서울, 인천, 광주 등 여러 지역에서 실업률을 설명하는 유의한 변수임을 알 수 있다.

<표 3-32> 25개 총별 실업률과 거주유형을 제외한 정렬기준변수와의 회귀분석

	모수	B	SE	t값	p값	R^2
서울	절편	0.0852	0.0173	4.91	<.0001	0.200
	광공업비율	-0.0427	0.0720	-0.59	0.5541	
	농림어업비율	-1.9841	0.7590	-2.61	0.0096	
	전월세비율	0.0380	0.0122	3.13	0.0020	
	30~40대비율	-0.1093	0.0480	-2.28	0.0236	
	30대 이상 대졸자비율	-0.0571	0.0221	-2.59	0.0102	
부산	절편	0.0348	0.0156	2.23	0.0275	0.044
	광공업비율	0.0144	0.0593	0.24	0.8089	
	농림어업비율	-0.2321	0.1860	-1.25	0.2148	
	전월세비율	0.0250	0.0143	1.74	0.0844	
	30~40대비율	-0.0388	0.0504	-0.77	0.4427	
	30대 이상 대졸자비율	0.0152	0.0261	0.58	0.5603	
대구	절편	0.0681	0.0218	3.13	0.0024	0.179
	광공업비율	0.0140	0.0808	0.17	0.8624	
	농림어업비율	-0.0566	0.0618	-0.92	0.3620	
	전월세비율	0.0407	0.0157	2.60	0.0110	
	30~40대비율	-0.1252	0.0785	-1.59	0.1143	
	30대 이상 대졸자비율	-0.0042	0.0473	-0.09	0.9290	
인천	절편	0.1090	0.0240	4.54	<.0001	0.196
	광공업비율	-0.1123	0.0958	-1.17	0.2435	
	농림어업비율	-0.2432	0.0619	-3.93	0.0001	
	전월세비율	-0.0046	0.0162	-0.29	0.7748	
	30~40대비율	-0.0348	0.0713	-0.49	0.6261	
	30대 이상 대졸자비율	-0.1517	0.0490	-3.09	0.0025	
광주	절편	0.0420	0.0189	2.22	0.0292	0.213
	광공업비율	-0.0785	0.0792	-0.99	0.3245	
	농림어업비율	0.0183	0.0628	0.29	0.7720	
	전월세비율	0.0212	0.0136	1.56	0.1223	
	30~40대비율	0.0312	0.0714	0.44	0.6634	
	30대 이상 대졸자비율	-0.0880	0.0406	-2.17	0.0333	
대전	절편	0.0595	0.0192	3.10	0.0027	0.177
	광공업비율	-0.0153	0.0957	-0.16	0.8731	
	농림어업비율	-0.0566	0.3677	-0.15	0.8781	
	전월세비율	0.0279	0.0141	1.98	0.0514	
	30~40대비율	-0.0722	0.0738	-0.98	0.3309	
	30대 이상 대졸자비율	-0.0320	0.0335	-0.95	0.3427	
울산	절편	0.1141	0.0264	4.32	<.0001	0.187
	광공업비율	0.0195	0.0584	0.33	0.7393	
	농림어업비율	-0.1587	0.0990	-1.60	0.1132	

	모수	B	SE	t값	p값	R^2
	전월세비용	0.0036	0.0185	0.19	0.8465	
	30~40대비용	-0.2043	0.0747	-2.73	0.0078	
	30대 이상 대졸자비용	-0.0135	0.0382	-0.35	0.7253	
경기동부	절편	0.0507	0.0174	2.91	0.0042	0.110
	광공업비용	-0.0381	0.0365	-1.05	0.2972	
	농림어업비용	-0.2030	0.1053	-1.93	0.0558	
	전월세비용	-0.0074	0.0108	-0.68	0.4983	
	30~40대비용	0.0366	0.0467	0.78	0.4347	
	30대 이상 대졸자비용	-0.0643	0.0175	-3.68	0.0003	
	절편	0.0569	0.0231	2.47	0.0166	
경기읍면부	광공업비용	0.0335	0.0429	0.78	0.4379	0.069
	농림어업비용	-0.0322	0.0330	-0.98	0.3334	
	전월세비용	-0.0154	0.0144	-1.07	0.2896	
	30~40대비용	-0.0620	0.0710	-0.87	0.3859	
	30대 이상 대졸자비용	-0.0246	0.0350	-0.70	0.4844	
	절편	0.0007	0.0396	0.02	0.9858	
강원동부	광공업비용	-0.2974	0.2554	-1.16	0.2508	0.063
	농림어업비용	0.0411	0.1664	0.25	0.8063	
	전월세비용	-0.0214	0.0210	-1.02	0.3125	
	30~40대비용	0.1724	0.1391	1.24	0.2222	
	30대 이상 대졸자비용	-0.0585	0.0655	-0.89	0.3772	
	절편	0.0473	0.0139	3.40	0.0015	
강원읍면부	광공업비용	0.2678	0.0748	3.58	0.0009	0.339
	농림어업비용	-0.0151	0.0155	-0.97	0.3360	
	전월세비용	-0.0107	0.0129	-0.82	0.4141	
	30~40대비용	-0.1078	0.0499	-2.16	0.0363	
	30대 이상 대졸자비용	-0.0147	0.0637	-0.23	0.8190	
	절편	0.0316	0.0163	1.94	0.0586	
충북동부	광공업비용	0.1005	0.0598	1.68	0.1002	0.236
	농림어업비용	0.0322	0.0573	0.56	0.5770	
	전월세비용	0.0082	0.0109	0.75	0.4559	
	30~40대비용	-0.0335	0.0593	-0.56	0.5754	
	30대 이상 대졸자비용	-0.0714	0.0309	-2.31	0.0257	
	절편	0.0116	0.0167	0.70	0.4890	
충북읍면부	광공업비용	0.0004	0.0381	0.01	0.9919	0.106
	농림어업비용	-0.0185	0.0168	-1.10	0.2785	
	전월세비용	-0.0111	0.0234	-0.48	0.6363	
	30~40대비용	0.0578	0.0592	0.98	0.3345	
	30대 이상 대졸자비용	-0.0661	0.0528	-1.25	0.2170	
	절편	0.0521	0.0264	1.98	0.0550	
충남동부	광공업비용	0.0455	0.0658	0.69	0.4937	0.103
	농림어업비용	-0.0496	0.1488	-0.33	0.7406	
	전월세비용	-0.0019	0.0161	-0.12	0.9069	
	30~40대비용	-0.0341	0.0868	-0.39	0.6961	
	30대 이상 대졸자비용	-0.0572	0.0409	-1.40	0.1702	
	절편	0.0589	0.0164	3.60	0.0008	
충남읍면부	광공업비용	0.1420	0.0393	3.62	0.0008	0.376
	농림어업비용	-0.0320	0.0173	-1.85	0.0707	
	전월세비용	-0.0103	0.0172	-0.60	0.5539	
	30~40대비용	-0.1842	0.0647	-2.85	0.0067	
	30대 이상 대졸자비용	0.1186	0.0470	2.52	0.0154	
	절편	0.0477	0.0178	2.67	0.0107	
전북동부	광공업비용	-0.0412	0.0827	-0.50	0.6215	0.044
	농림어업비용	-0.0297	0.0890	-0.33	0.7405	

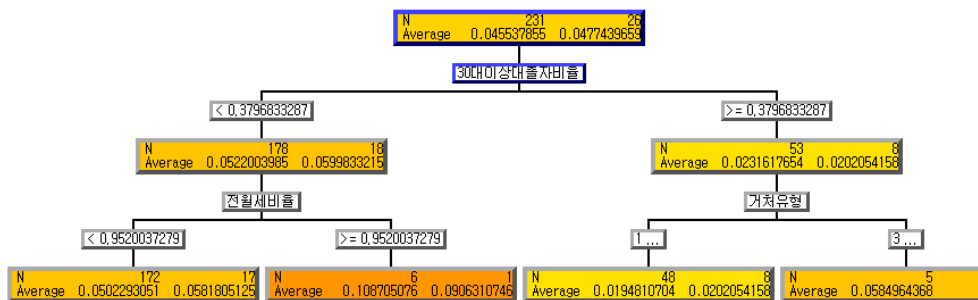
	모수	B	SE	t값	p값	R^2
	전월세비용	-0.0028	0.0114	-0.24	0.8081	
	30~40대비용	-0.0412	0.0601	-0.68	0.4972	
	30대 이상 대졸자비용	-0.0111	0.0353	-0.31	0.7543	
전북읍면부	절편	-0.0095	0.0179	-0.53	0.6001	0.068
	광공업비용	0.0591	0.1004	0.59	0.5603	
	농림어업비용	0.0219	0.0226	0.97	0.3379	
	전월세비용	0.0197	0.0276	0.71	0.4810	
	30~40대비용	0.0643	0.0778	0.83	0.4140	
	30대 이상 대졸자비용	-0.0872	0.0818	-1.07	0.2935	
	절편	0.0887	0.0217	4.09	0.0002	
전남동부	광공업비용	0.0173	0.0551	0.31	0.7544	0.173
	농림어업비용	-0.1281	0.1151	-1.11	0.2720	
	전월세비용	-0.0061	0.0126	-0.48	0.6318	
	30~40대비용	-0.1674	0.0716	-2.34	0.0242	
	30대 이상 대졸자비용	0.0070	0.0413	0.17	0.8656	
전남읍면부	절편	0.0480	0.0162	2.96	0.0050	0.326
	광공업비용	-0.0191	0.0760	-0.25	0.8029	
	농림어업비용	-0.0424	0.0167	-2.54	0.0149	
	전월세비용	0.0315	0.0184	1.71	0.0943	
	30~40대비용	-0.0847	0.0537	-1.58	0.1221	
	30대 이상 대졸자비용	0.0063	0.0505	0.12	0.9016	
경북동부	절편	0.0340	0.0230	1.48	0.1442	0.091
	광공업비용	0.0024	0.0340	0.07	0.9433	
	농림어업비용	-0.0292	0.0559	-0.52	0.6034	
	전월세비용	0.0028	0.0142	0.20	0.8443	
	30~40대비용	0.0218	0.0659	0.33	0.7426	
경북읍면부	30대 이상 대졸자비용	-0.0775	0.0432	-1.80	0.0783	0.263
	절편	0.0061	0.0127	0.48	0.6317	
	광공업비용	0.0596	0.0450	1.32	0.1907	
	농림어업비용	-0.0018	0.0159	-0.11	0.9095	
	전월세비용	0.0221	0.0187	1.18	0.2433	
	30~40대비용	0.0077	0.0409	0.19	0.8520	
경남동부	30대 이상 대졸자비용	0.0323	0.0402	0.80	0.4247	0.143
	절편	0.0209	0.0292	0.72	0.4774	
	광공업비용	-0.0114	0.0715	-0.16	0.8739	
	농림어업비용	0.1215	0.1548	0.79	0.4356	
	전월세비용	0.0202	0.0181	1.12	0.2688	
	30~40대비용	0.0541	0.0958	0.56	0.5744	
경남읍면부	30대 이상 대졸자비용	-0.0910	0.0444	-2.05	0.0452	0.236
	절편	-0.0016	0.0171	-0.09	0.9253	
	광공업비용	0.0375	0.0364	1.03	0.3078	
	농림어업비용	0.0001	0.0214	0.01	0.9960	
	전월세비용	0.0313	0.0171	1.82	0.0743	
	30~40대비용	0.0373	0.0510	0.73	0.4684	
제주동부	30대 이상 대졸자비용	-0.0361	0.0434	-0.83	0.4100	0.257
	절편	-0.0362	0.0213	-1.70	0.1023	
	광공업비용	0.1383	0.1815	0.76	0.4533	
	농림어업비용	0.0239	0.0318	0.75	0.4597	
	전월세비용	0.0038	0.0105	0.36	0.7220	
	30~40대비용	0.1558	0.0684	2.28	0.0320	
제주읍면부	30대 이상 대졸자비용	-0.0323	0.0299	-1.08	0.2907	0.460
	절편	0.0672	0.0237	2.84	0.0132	
	광공업비용	0.0124	0.2202	0.06	0.9558	
	농림어업비용	-0.0232	0.0219	-1.06	0.3072	

	모수	B	SE	t값	p값	R ²
	전월세비율	0.0311	0.0266	1.17	0.2632	
	30~40대비율	-0.2151	0.0793	-2.71	0.0168	
	30대 이상 대졸자비율	0.1096	0.0568	1.93	0.0743	

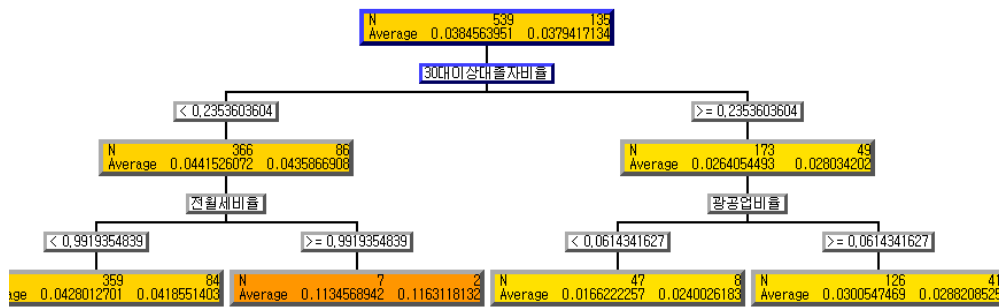
(3) 정렬변수와 실업률의 의사결정나무분석

실업률을 목표변수로 하여 기존의 정렬변수들과 추가 정렬기준 변수들 중에서 각 층별로 분류 기준변수가 되는 변수들을 순차적으로 파악해 보기 위해, 우선 전체 조사구를 4개 지역(서울, 광역시, 동부, 읍면부)으로 나눈 후에 의사결정나무분석을 실시하였다. 본 연구에서는 25개 층별로 의사결정나무 분석도 실시하였으며, 25개 층별 의사결정나무 분석결과의 신뢰도를 높이기 위해 특정 층 내의 조사구 수가 30이상인 층에서만 의사결정나무분석을 실시하였고, 이 기준에 의해 제주읍면부는 분석에서 제외되었다.

서울, 광역시, 동부, 읍면부등 4개 지역으로 나누었을 때의 의사결정나무분석결과는 [그림 3-1]~[그림 3-4]와 같다.



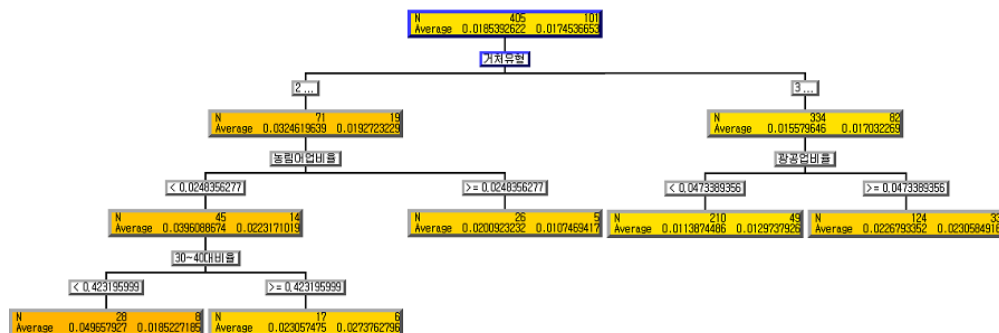
[그림 3-1] 실업률-서울



[그림 3-2] 실업률-광역시



[그림 3-3] 실업률-동부



[그림 3-4] 실업률-읍면부

서울의 실업률은 30대 이상 대졸자비율이 첫 번째 분류기준 변수였으며, 약37.9% 이상인 것과 37.9%미만인 것으로 분류가 되며, 30대 이상 대졸자비율이 37.9%이상인 경우에는 실업률의 평균이 약2.3%이고, 30대 이상 대졸자비율이 37.9%미만인 경우에는 실업률의 평균이 약5.22%였다.

광역시의 실업률 역시 30대 이상 대졸자비율이 첫 번째 분류기준 변수였으며, 약

23.5%이상인 것과 23.5%미만인 것으로 분류가 되며, 30대 이상 대졸자비율이 37.9% 이상인 경우에는 실업률의 평균이 약 2.64%이고, 30대 이상 대졸자비율이 37.9%미만인 경우에는 실업률의 평균이 약 4.41%이다.

동부의 경우 실업률은 거처유형이 첫 번째 분류기준변수이며 거처유형이 4(기타: 아파트와 단독주택을 제외한 유형)인 곳과 그 외 유형으로 분류가 되며, 거처유형이 4인 경우에는 실업률의 평균이 약 4.92%이고, 나머지 거처유형에서는 실업률의 평균이 약 2.9%이다.

읍면부의 실업률도 거처유형이 첫 번째 분류기준변수이며, 거처유형이 2, 4(중대형/대형아파트, 기타)인 곳과 거처유형이 1, 3(소형/중소형아파트, 단독주택인 곳으로 분류가 되며, 거처유형이 2, 4인 경우에는 실업률의 평균이 약 3.2%이고, 거처유형이 1, 3인 경우에는 실업률의 평균이 약 1.5%이다.

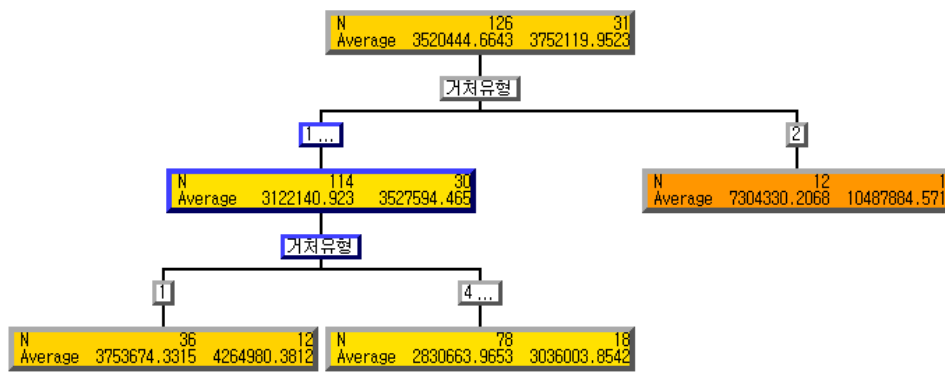
분석된 결과를 종합적으로 정리해 보면, 서울과 광역시 등과 같은 도시 지역의 첫 번째로 선택된(중요한) 분류기준 변수는 30대 이상 대졸자비율이었고, 도지역의 동부와 읍면부의 경우 첫 번째 분류기준변수는 거처유형인 것으로 나타났다. 이런 분석 결과는 향후 표본설계에 반영함으로써 표본설계의 효율성을 제공할 수 있는 가치 있는 정보를 제공하고 있다고 보인다.

한편 목표변수가 가계동향조사의 경상소득인 경우 4개 지역 구분에서의 의사결정 나무분석 분석결과는 [그림 3-5]~[그림 3-8]와 같다. 서울의 경우 경상소득은 거처유형이 2(중대형/대형아파트)인 곳과 그 외 지역으로 분류가 되며, 거처유형이 2인 곳의 경상소득의 평균은 7,304,330원이고 그 외 지역의 경상소득의 평균은 3,122,140원이다.

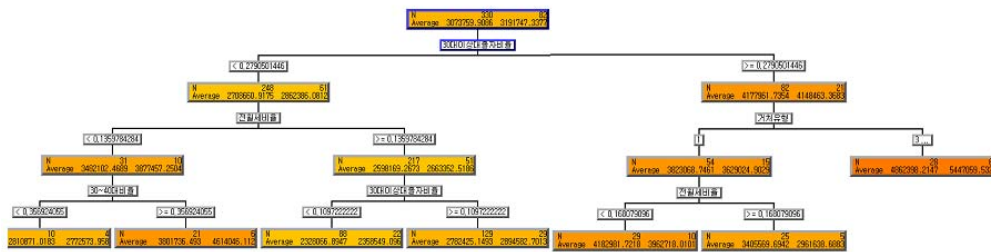
광역시의 경우 경상소득은 30대 이상 대졸자비율이 27.9% 이상인 곳과 27.9% 미만인 곳으로 분류가 되며, 30대 이상 대졸자비율이 27.9% 이상인 곳의 경상소득의 평균은 4,177,961원이고, 30대 이상 대졸자비율이 27.9% 미만인 곳의 경상소득의 평균은 2,708,660원이다.

동부의 경우 경상소득은 30대 이상 대졸자비율이 25.6% 이상인 곳과 25.6% 미만인 곳으로 분류가 되며, 30대 이상 대졸자비율이 25.6% 이상인 곳의 경상소득의 평균은 4,051,206원이고, 30대 이상 대졸자비율이 25.6% 미만인 곳의 경상소득의 평균은 2,676,446원이다.

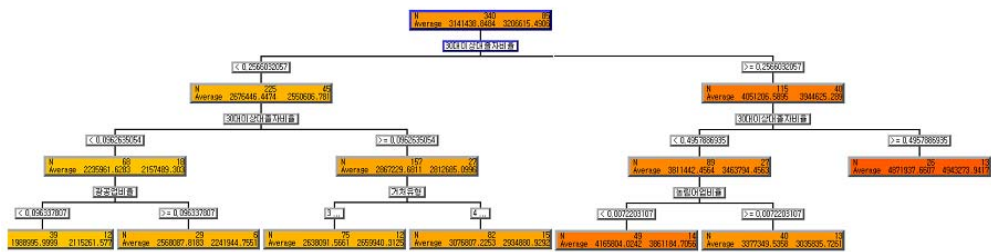
읍면부의 경우 경상소득은 30~40대 비율이 27.8%이상인 곳과 27.8% 미만인 곳으로 분류되었으며, 30~40대 비율이 27.8% 이상인 곳의 경상소득의 평균은 2,970,667원이고, 30~40대 비율이 27.8% 이상인 곳의 경상소득의 평균은 1,845,324원이다.



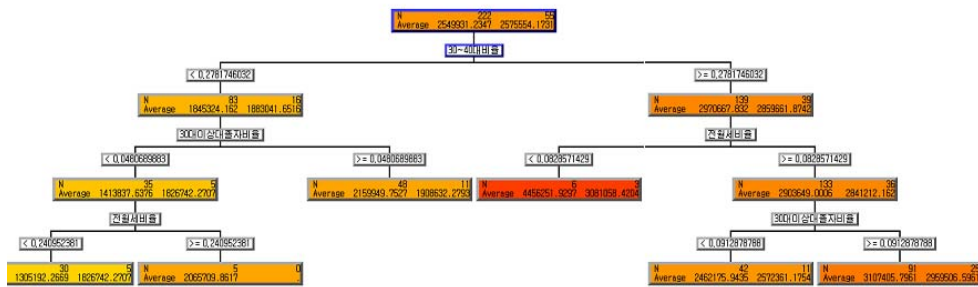
[그림 3-5] 경상소득-서울



[그림 3-6] 경상소득-광역시



[그림 3-7] 경상소득-동부



[그림 3-8] 경상소득-읍면부

다음으로, 25개 층별 의사결정나무 분석을 한 결과 전체 25개 층 중에서 첫 번째 분류기준 변수가 30대 이상 대졸자비율인 층이 서울 등 6개 층으로 가장 많았고, 전월세비율이 첫 번째 분류기준 변수인 층이 5개 층, 30~40대 비율이 3개 층, 농림어업비율이 3개 층, 거취유형 3개 층, 광공업비율이 2개 층인 것으로 나타났다. 한편 부산, 강원동부에 대한 의사결정나무는 잘 적합이 되지 않았고, 층 내 조사구수가 30미만인 층은 분석에서 제외함에 따라 제주읍면부 층은 분석대상에서 제외되었다.

이 들 세 개 층(부산, 강원 동부, 제주 읍면부)을 제외한 나머지 22개 층에 대한 의사결정나무 분석결과를 요약한 결과는 <표 3-33>과 같다.

〈표 3-33〉 층별 분류기준변수

첫 번째 분류기준변수	층	두 번째 분류기준변수
30대 이상 대졸자비율	서울	전월세비율, 거처유형
	경기읍면부	없음
	충남동부	농림어업비율
	전북동부	농림어업비율
	전남동부	없음
	경북동부	광공업비율
전월세비율	대구	거처유형
	광주	30대 이상 대졸자비율
	전북읍면부	30대 이상 대졸자비율, 농림어업비율
	전남읍면부	농림어업비율
	제주동부	광공업비율
30~40대비율	대전	30대 이상 대졸자비율
	울산	농림어업비율, 광공업비율
	강원읍면부	없음
거처유형	경기동부	광공업비율
	경남동부	30대 이상 대졸자비율
	경남읍면부	30~40대비율
농림어업비율	인천	30대 이상 대졸자비율
	충북읍면부	없음
	충남읍면부	30대 이상 대졸자비율
광공업비율	충북동부	30대 이상 대졸자비율
	경북읍면부	30대 이상 대졸자비율, 전월세비율

2) 표본 크기

(1) 최적 표본 가구수 분석

현행 경제활동인구조사와 가계동향조사는 25개 지역(층)별로 조사구를 추출하고 각각의 조사구에서 경활은 20가구, 가계는 10가구를 대상으로 조사를 수행한다. 향후 표본설계에서도 현재와 같은 조사구 내 표본 가구수를 그대로 유지하는 것이 타당한지에 대한 검토가 필요하며, 만약 조사구 내 표본 가구수에 변화가 필요한 경우 표본설계를 상당 부분 수정할 필요가 있다. 이런 측면에서 여기서는 조사구 내 최적

표본 가구수를 이론적으로 산출해 보기 위해 2010년 1월부터 12월까지의 데이터를 기준으로 다음과 같은 분석을 수행하였다.

한 가지 유의할 점은 경제활동인구조사는 조사구에서 가구를 추출한 후에 표본 가구의 15세 이상 모든 가구원을 대상으로 조사가 이루어지므로 이론적으로 가구 기준 집락규모($m=20$)의 적절성을 검토하기 위해서는 우선 가구원 기준으로 조사된 자료를 가구단위로 변환해야 한다. 따라서 본 연구에서는 이론적으로 타당한 집락효과 분석을 위해 주요 변수별로 가구원 합계 또는 평균을 계산하여 가구 기준 변수를 생성하였고, 조사구 및 가구 조사 비용의 상대적인 비율을 고려하여 변수별로 조사구 내 최적 가구수를 산출한 후 이를 종합하는 방식으로 최종 결론을 얻을 것이라는 점을 밝혀둔다.

가구 기준 변수 생성 후 최적 표본 가구수 산출 과정을 정리하면, 각 지역(층) 및 월별 분산분석 결과를 이용하여 평균 집락효과(Deff)와 R_a^2 , MSw, MSb를 산출하였고, 조사구당 조사비용(C_{psu})과 조사구내 가구당 조사비용(C_{ssu})의 상대적인 비율($C_{psu}:C_{ssu}$)을 가정한 후, 최적 가구수를 산정하였다. 조사구내 최적 가구수를 구하기 위해 사용한 공식은 다음과 같다.

$$m_{opt} = \sqrt{\frac{s_w^2 c_{psu}}{s_b^2 c_{ssu}}}$$

$$MSw = \frac{1}{n(m-1)} \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2, \quad MSb = \frac{m}{n-1} \sum_{i=1}^n (\bar{y}_i - \bar{\bar{y}})^2$$

$$S_w^2 = MSw, \quad S_b^2 = \frac{MSb - MSw}{m}, \quad S^2 = \frac{1}{nm-1} \sum_{i=1}^n \sum_{j=1}^m (y_{ij} - \bar{\bar{y}})^2$$

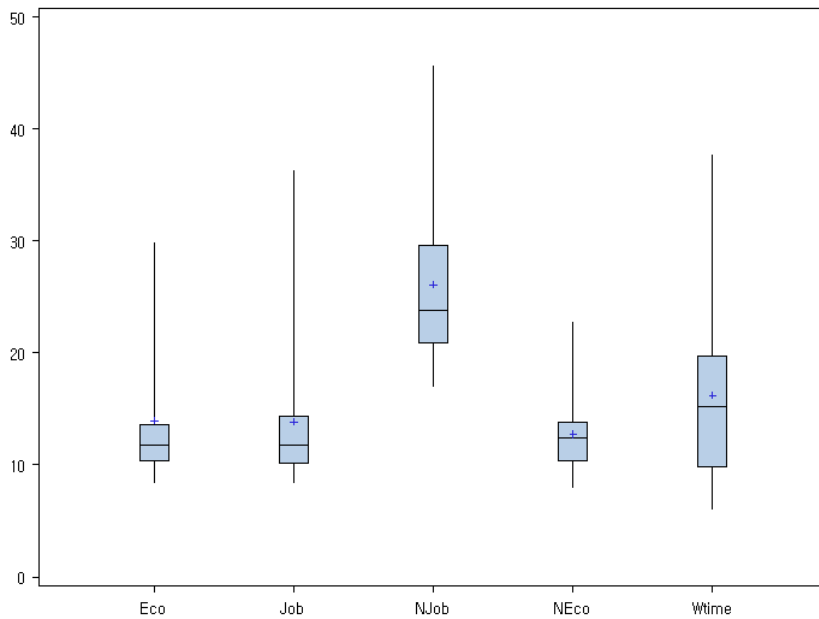
$$deff = \frac{MSb}{S^2}, \quad R_a^2 = 1 - \frac{MSw}{S^2}$$

최적 표본 가구수는 경활인구, 취업인구, 실업인구, 비경활인구, 취업시간을 기준으로 산출하였으며, 최종 최적 가구수로는 변수별 최적 가구수의 중앙값을 사용하였다. 조사비용 비율에 따른 조사구 당 최적 표본가구수 산출결과는 <표 3-34>와 같다.

<표 3-34>를 보면, 조사구 대비 가구 조사비용이 5:1일 때 (일반적으로 통계청에서 생각하는 조사구 대비 가구 조사비용에 해당함) 평균 최적 표본 가구수는 15가구이며, 비용에 따라 9~21가구가 적절한 것으로 나타났다. 대체적으로 현재 경제활동 인구조사에서의 조사구 당 표본 가구수인 20가구에 비해 일정 부분 표본을 줄이는 것을 고려할 필요가 있다는 점을 시사하고 있다. 그러나 총별로 차이가 있고 도지역의 경우 동부 지역이 읍면부 지역보다 상대적으로 조사구 당 표본 가구수를 크게 하는 것이 효율적임을 볼 수 있다.

<표 3-34> 경제활동인구조사의 총별 조사구당 최적 표본 가구수

$C_{psu} \cdot C_{ssu}$ 총	2:1	4:1	5:1	6:1	8:1	10:1
평균	9	13	15	16	18	21
서울	10	14	16	17	20	22
부산	8	11	12	13	15	17
대구	7	10	11	12	14	15
인천	7	10	12	13	15	16
광주	8	11	13	14	16	18
대전	8	12	13	15	17	19
울산	9	12	14	15	18	20
경기/동부	13	18	20	22	25	28
경기/읍면	9	12	14	15	18	20
강원/동부	10	15	16	18	21	23
강원/읍면	8	11	12	13	15	17
충북/동부	8	11	13	14	16	18
충북/읍면	6	9	10	11	13	14
충남/동부	9	13	14	16	18	20
충남/읍면	7	11	12	13	15	17
전북/동부	13	19	21	23	26	29
전북/읍면	8	12	13	14	16	18
전남/동부	9	12	14	15	17	19
전남/읍면	6	8	9	10	12	13
경북/동부	12	16	18	20	23	26
경북/읍면	8	11	12	13	15	17
경남/동부	8	12	13	14	17	19
경남/읍면	8	12	13	15	17	19
제주/동부	12	18	20	21	25	28
제주/읍면	19	27	30	33	38	42



[그림 3-9] 경제활동인구조사의 변수별 최적 표본 가구수 분포(5:1)

[그림 3-9]는 조사비용의 비율이 5:1일 때 변수별로 층별 최적 표본 가구수의 분포를 요약한 것이다. 변수에 따라 층별 변동이 크지만 실업인구(Njob)를 제외한 다른 변수는 15가구 이내를 중심으로 분포하고 있다는 것을 볼 수 있다. 분석결과를 종합하면 향후 경제활동인구조사 표본설계에서는 기존의 20가구 대신 15가구 수준으로 조사구당 표본 가구수를 조정하는 방안을 적극적으로 검토할 필요가 있다고 판단된다.

<표 3-34>는 경제활동인구조사 변수 중 경활인구, 취업인구, 실업인구, 비경활인구, 취업시간에 대한 월별 자료에 대한 분석결과를 모두 종합해 얻은 결과라는 점에 유의하기 바란다. 예를 들어 실업인구를 대상으로 하는 경우 조사구당 최적 표본 가구수에 대한 분석결과는 <표 3-35>와 같다는 점을 참고하기 바란다. [그림 3-9]는 변수별로 <표 3-35>와 같은 방식으로 분석된 결과들을 요약해 그린 상자그림이라는 점을 다시 한 번 밝혀둔다. 다른 변수들에 대한 최적 표본 가구수 산출 결과는 부록에 수록하였다.

〈표 3-35〉 〈경찰〉 실업인구의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$)별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	1.36	0.023	11	15	17	19	21	24
부산	1.17	0.010	15	21	24	26	30	34
대구	1.18	0.011	20	28	32	35	40	45
인천	1.17	0.011	17	24	26	29	33	37
광주	1.26	0.017	13	19	21	23	27	30
대전	1.29	0.017	12	17	19	21	25	27
울산	1.16	0.010	13	18	21	23	26	29
경기/동부	1.12	0.007	19	26	30	32	37	42
경기/읍면	1.06	0.004	29	41	46	50	58	65
강원/동부	1.19	0.011	13	19	21	23	27	30
강원/읍면	1.27	0.016	15	22	24	27	31	34
충북/동부	1.21	0.013	14	19	21	24	27	30
충북/읍면	1.14	0.008	15	21	24	26	30	34
충남/동부	1.24	0.015	19	27	30	33	38	43
충남/읍면	1.12	0.007	19	26	30	32	37	42
전북/동부	1.18	0.011	13	19	21	23	26	29
전북/읍면	1.16	0.009	24	34	38	42	49	54
전남/동부	1.14	0.008	15	21	24	26	30	33
전남/읍면	1.11	0.006	18	25	28	31	36	40
경북/동부	1.30	0.018	14	20	23	25	29	32
경북/읍면	1.30	0.018	16	23	26	28	33	37
경남/동부	1.41	0.026	11	15	17	19	21	24
경남/읍면	1.12	0.007	24	34	38	41	48	53
제주/동부	1.20	0.012	12	18	20	21	25	28
제주/읍면	1.12	0.007	20	28	31	34	40	44

가계동향조사 자료를 대상으로 경제활동인구조사와 마찬가지로 집락규모($m=10$)의 적절성을 검토하기 위해 조사구 대비 가구의 조사비용의 상대적인 비율을 고려하여 최적 가구수를 산출하였다. 가계동향조사에서 최적 표본가구수 산출을 위해 분석한 변수는 소득(A), 경상소득(B2), 근로소득(B4), 가계지출(D), 소비지출(F)이며, <표 3-36>의 최종 최적 가구수는 변수별로 얻은 최적 가구수의 중앙값에 해당한다.

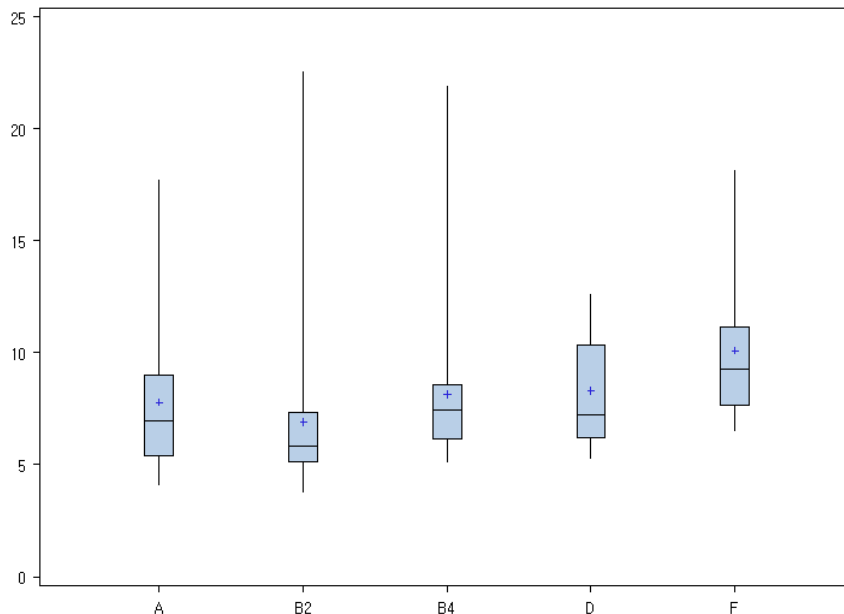
<표 3-36> 가계동향조사의 층별 조사구당 최적 표본 가구수

$C_{psu} \cdot C_{ssu}$ 층	2:1	4:1	5:1	6:1	8:1	10:1
평균	5	7	8	9	10	11
서울	4	5	6	6	7	8
부산	8	11	12	13	15	17
대구	4	5	6	6	7	8
인천	5	8	9	10	11	12
광주	4	6	6	7	8	9
대전	4	6	6	7	8	9
울산	7	9	11	12	13	15
경기/동부	4	6	6	7	8	9
경기/읍면	6	8	9	10	12	13
강원/동부	7	10	11	12	14	16
강원/읍면	4	6	7	7	9	10
충북/동부	5	7	7	8	9	11
충북/읍면	5	7	8	8	10	11
충남/동부	4	6	7	7	8	9
충남/읍면	4	6	6	7	8	9
전북/동부	6	8	9	10	11	12
전북/읍면	3	5	5	6	7	7
전남/동부	5	7	8	8	10	11
전남/읍면	4	6	7	7	8	9
경북/동부	3	5	5	6	7	8
경북/읍면	4	5	6	6	7	8
경남/동부	4	6	7	7	8	9
경남/읍면	5	7	7	8	9	10
제주/동부	5	7	8	9	10	11
제주/읍면	10	14	16	18	20	23

경제활동인구조사와 마찬가지로 조사비용의 비율이 5:1인 경우를 보면, 평균 최적 표본 가구수는 8가구이며, 비용에 따라 5~11가구가 적절한 것으로 나타났다. 부산과

강원 동부, 제주 읍면부 지역을 제외하고는 층별 차이는 크지 않다. 따라서 가계동향조사의 경우 현재와 마찬가지로 조사구 당 10가구를 조사하는 방식을 그대로 유지하는 것이 바람직한 것으로 판단된다.

한편 [그림 3-10]은 조사비용의 비율이 5:1일 때 가계동향조사 분석대상 변수별로 층별 최적 표본 가구수의 분포를 나타낸 것이다. 총소득(A)과 경상소득(B2), 근로소득(B4)의 층별 변동이 크지만 대체로 10가구를 약간 하회하는 수준을 중심으로 분포하고 있다는 것을 볼 수 있다.



[그림 3-10] 가계동향조사의 변수별 최적 표본 가구수 분포(5:1)

<표 3-36>은 가계동향조사 변수 중 소득, 경상소득, 근로소득, 가계지출, 소비지출에 대한 월별 자료 분석결과를 모두 종합해 얻은 결과라는 점에 유의하기 바라고, 예를 들어, 소비지출을 기준으로 매월 층별 $deff$ 와 R_b^2 , 조사비용의 비율에 따른 표본 가구수를 산출한 후에 12개월 동안의 평균값을 구하여 최적 표본 가구수를 산출한 결과는 <표 3-37>과 같다. 다른 변수들에 대한 최적 표본 가구수 산출 결과는 부록에 수록하였다.

〈표 3-37〉 〈가계〉 소비지출의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$) 별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	1.75	0.131	5	7	7	8	9	10
부산	1.38	0.064	8	11	12	13	15	17
대구	2.00	0.154	4	6	6	7	8	9
인천	1.59	0.095	6	8	9	10	11	13
광주	1.71	0.122	5	6	7	8	9	10
대전	1.63	0.099	5	7	8	9	10	11
울산	1.30	0.048	11	15	17	18	21	24
경기/동부	1.62	0.096	6	8	9	10	12	13
경기/읍면	1.25	0.043	11	16	18	20	23	26
강원/동부	1.37	0.058	10	14	15	17	20	22
강원/읍면	1.52	0.078	6	9	10	11	13	14
충북/동부	1.45	0.070	8	11	12	14	16	17
충북/읍면	1.40	0.073	7	10	11	13	15	16
충남/동부	1.55	0.091	5	7	8	9	10	11
충남/읍면	1.55	0.086	6	9	10	10	12	14
전북/동부	1.49	0.077	6	8	9	10	12	13
전북/읍면	2.24	0.223	4	6	7	7	9	10
전남/동부	1.62	0.096	6	8	9	10	12	13
전남/읍면	1.84	0.143	5	7	8	8	10	11
경북/동부	1.73	0.110	5	7	8	9	10	11
경북/읍면	1.88	0.157	4	6	7	8	9	10
경남/동부	1.87	0.145	5	7	8	9	10	12
경남/읍면	1.36	0.058	6	8	9	10	12	13
제주/동부	1.44	0.073	6	9	10	10	12	14
제주/읍면	1.04	0.008	10	14	16	18	20	23

(2) 주요 관심 변수의 상대표준오차

향후 표본설계의 표본크기를 가늠하기 위해 현행 2010년 경제활동인구조사 및 가계동향조사 자료를 이용하여 주요 변수에 대한 상대표준오차(CV)를 계산하였다. 각 변수별로 매월 전국 및 시도, 층별 상대표본오차와 함께 월별 상대표준오차의 중앙값도 함께 정리하였다. 경제활동인구조사의 실업인구 총계 추정에 대한 시도별 상대표준오차와 층별 상대표준오차 산출 결과는 <표 3-38>과 <표 3-39>와 같다.

제시된 상대표준오차는 이론적인 측면에서 보면 경제활동인구조사에서 추출단위가 가구이라는 점을 감안해, 앞 절에서 생성한 가구원 변수를 종합해 생성한 가구단위 실업인구 변수의 상대표준오차를 산출하였다는 점에 유의하기 바란다. 분석 결과 상대표준오차는 전국은 3.2%, 서울, 경기도는 8%, 광역시는 12%내외인 것으로 나타났으며, 제주를 제외한 도 지역은 15%내외로 원래 표본설계 당시에 설정한 목표 정도보다 조금 크게 나타났으며, 월별로는 큰 차이를 보이지 않고 있다.

참고로 통계청에서 가구부문 3차 연동표본의 표본규모를 산정할 때, 경제활동인구조사의 경우 실업인구의 연평균 상대표준오차를 기준으로 전국의 경우 목표 정도를 3%이하, 시도 단위에서는 서울과 경기도를 6%내외, 광역시는 10%내외, 기타 도는 14%내외로 설정하였다.

〈표 3-38〉 〈경찰〉 실업인구 총계의 시도 및 월별 상대표준오차

	중앙값	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	0.032	0.028	0.029	0.031	0.031	0.034	0.033	0.033	0.034	0.032	0.032	0.035	0.033
서울	0.079	0.073	0.073	0.080	0.078	0.089	0.088	0.080	0.088	0.076	0.073	0.079	0.073
부산	0.120	0.094	0.098	0.103	0.104	0.120	0.113	0.120	0.126	0.125	0.132	0.139	0.140
대구	0.125	0.106	0.120	0.129	0.125	0.125	0.128	0.126	0.146	0.135	0.123	0.113	0.101
인천	0.098	0.097	0.084	0.098	0.095	0.110	0.102	0.103	0.104	0.099	0.096	0.110	0.095
광주	0.148	0.113	0.139	0.149	0.138	0.132	0.140	0.182	0.146	0.162	0.180	0.167	0.161
대전	0.137	0.129	0.107	0.138	0.130	0.151	0.131	0.121	0.136	0.155	0.149	0.161	0.158
울산	0.143	0.112	0.135	0.136	0.131	0.159	0.137	0.154	0.164	0.147	0.142	0.158	0.145
경기	0.080	0.076	0.068	0.070	0.075	0.085	0.078	0.082	0.084	0.079	0.083	0.093	0.083
강원	0.174	0.117	0.164	0.176	0.155	0.174	0.189	0.196	0.221	0.175	0.171	0.201	0.173
충북	0.176	0.150	0.160	0.170	0.172	0.189	0.208	0.187	0.193	0.167	0.180	0.161	0.187
충남	0.153	0.121	0.141	0.158	0.157	0.147	0.153	0.145	0.153	0.141	0.155	0.170	0.170
전북	0.182	0.157	0.152	0.174	0.195	0.185	0.170	0.182	0.191	0.170	0.182	0.249	0.198
전남	0.164	0.140	0.168	0.188	0.182	0.195	0.196	0.161	0.148	0.156	0.135	0.148	0.187
경북	0.150	0.122	0.127	0.149	0.139	0.147	0.148	0.168	0.161	0.153	0.153	0.151	0.160
경남	0.147	0.134	0.141	0.129	0.137	0.155	0.147	0.157	0.157	0.140	0.148	0.201	0.173
제주	0.237	0.198	0.223	0.219	0.210	0.236	0.234	0.269	0.259	0.238	0.291	0.276	0.293

〈표 3-39〉 〈경찰〉 실업인구 총계의 총 및 월별 상대표준오차

	중앙값	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	0.032	0.028	0.029	0.031	0.031	0.034	0.033	0.033	0.034	0.032	0.032	0.035	0.033
서울	0.079	0.073	0.073	0.080	0.078	0.089	0.088	0.080	0.088	0.076	0.073	0.079	0.073
부산	0.120	0.094	0.098	0.103	0.104	0.120	0.113	0.120	0.126	0.125	0.132	0.139	0.140
대구	0.125	0.106	0.120	0.129	0.125	0.125	0.128	0.126	0.146	0.135	0.123	0.113	0.101
인천	0.098	0.097	0.084	0.098	0.095	0.110	0.102	0.103	0.104	0.099	0.096	0.110	0.095
광주	0.148	0.113	0.139	0.149	0.138	0.132	0.140	0.182	0.146	0.162	0.180	0.167	0.161
대전	0.137	0.129	0.107	0.138	0.130	0.151	0.131	0.121	0.136	0.155	0.149	0.161	0.158
울산	0.143	0.112	0.135	0.136	0.131	0.159	0.137	0.154	0.164	0.147	0.142	0.158	0.145
경기/동부	0.093	0.091	0.080	0.082	0.083	0.098	0.090	0.095	0.099	0.089	0.097	0.113	0.097
경기/읍면	0.161	0.136	0.132	0.136	0.175	0.169	0.149	0.161	0.163	0.167	0.162	0.161	0.157
강원/동부	0.218	0.184	0.194	0.197	0.200	0.216	0.235	0.253	0.281	0.219	0.197	0.255	0.230
강원/읍면	0.298	0.152	0.274	0.307	0.242	0.289	0.319	0.307	0.347	0.289	0.350	0.319	0.256
충북/동부	0.225	0.191	0.208	0.226	0.223	0.214	0.274	0.264	0.243	0.193	0.230	0.199	0.253
충북/읍면	0.270	0.243	0.236	0.256	0.261	0.386	0.304	0.260	0.320	0.269	0.281	0.271	0.270
충남/동부	0.196	0.151	0.169	0.195	0.191	0.192	0.208	0.197	0.216	0.186	0.220	0.228	0.226
충남/읍면	0.222	0.202	0.218	0.264	0.273	0.226	0.227	0.214	0.212	0.217	0.208	0.246	0.243
전북/동부	0.199	0.165	0.197	0.188	0.216	0.198	0.192	0.221	0.232	0.185	0.200	0.310	0.261
전북/읍면	0.385	0.348	0.238	0.435	0.447	0.452	0.367	0.312	0.333	0.424	0.405	0.403	0.295
전남/동부	0.208	0.190	0.214	0.244	0.213	0.226	0.255	0.203	0.186	0.190	0.149	0.171	0.241
전남/읍면	0.283	0.201	0.261	0.289	0.348	0.378	0.281	0.256	0.242	0.271	0.305	0.287	0.286
경북/동부	0.189	0.130	0.169	0.187	0.190	0.181	0.182	0.195	0.197	0.192	0.188	0.191	0.204
경북/읍면	0.248	0.230	0.193	0.234	0.196	0.250	0.253	0.330	0.272	0.253	0.264	0.239	0.246
경남/동부	0.185	0.171	0.165	0.152	0.158	0.192	0.191	0.219	0.210	0.172	0.178	0.249	0.215
경남/읍면	0.247	0.215	0.259	0.240	0.268	0.259	0.226	0.213	0.219	0.227	0.254	0.327	0.261
제주/동부	0.292	0.267	0.284	0.230	0.227	0.299	0.277	0.321	0.328	0.267	0.390	0.332	0.386
제주/읍면	0.426	0.279	0.341	0.585	0.542	0.375	0.426	0.492	0.423	0.490	0.426	0.492	0.420

한편, 가계동향조사의 경우 통계청 표본설계에서 기준 변수로 사용했던 소비지출 변수에 대한 상대표준오차를 산출한 결과는 <표 3-40> 및 <표 3-41>과 같다. 현행 2010년 가계동향조사에 대한 상대표준오차 분석 결과 전국은 1.7%, 서울, 경기 4% 내외, 제주를 제외한 광역시 및 기타 도 지역은 7%내외로 나타났으며 월별로는 큰 차이를 보이지 않고 있다. 가계동향조사의 상대표준오차를 동부와 읍면부로 나눠서 살펴보면, 전국, 도시의 목표정도를 2%이하, 읍면부는 4%이하로 목표정도를 만족하고 있는 것으로 나타났다.

참고로 통계청에서 3차 연동표본의 표본규모를 산정할 때, 가계동향조사는 소비지출의 연평균 상대표준오차에 대한 목표정도를 전국, 도시의 경우 2%이하, 읍면부는 5%이하로 설정하였다.

<표 3-40> <가계> 소비지출 평균의 시도 및 월별 상대표준오차

	중앙값	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	0.017	0.018	0.019	0.017	0.017	0.016	0.019	0.017	0.018	0.016	0.016	0.016	0.018
서울	0.042	0.042	0.046	0.042	0.041	0.038	0.042	0.042	0.041	0.042	0.038	0.039	0.043
부산	0.047	0.051	0.055	0.047	0.050	0.043	0.045	0.041	0.053	0.046	0.041	0.044	0.049
대구	0.053	0.053	0.059	0.053	0.051	0.051	0.055	0.059	0.058	0.053	0.050	0.049	0.048
인천	0.054	0.053	0.064	0.064	0.043	0.045	0.070	0.045	0.058	0.054	0.038	0.041	0.065
광주	0.055	0.055	0.076	0.053	0.088	0.047	0.055	0.052	0.061	0.048	0.047	0.060	0.059
대전	0.058	0.057	0.075	0.056	0.064	0.055	0.058	0.077	0.058	0.054	0.067	0.054	0.060
울산	0.072	0.085	0.087	0.066	0.086	0.097	0.087	0.063	0.070	0.051	0.051	0.073	0.066
경기	0.043	0.047	0.051	0.040	0.043	0.040	0.053	0.040	0.049	0.041	0.042	0.044	0.044
강원	0.049	0.051	0.068	0.056	0.045	0.043	0.074	0.074	0.047	0.036	0.045	0.045	0.054
충북	0.065	0.075	0.083	0.079	0.059	0.060	0.087	0.102	0.063	0.059	0.053	0.058	0.067
충남	0.070	0.097	0.061	0.061	0.072	0.060	0.060	0.069	0.071	0.073	0.080	0.060	0.104
전북	0.066	0.061	0.093	0.064	0.098	0.067	0.066	0.074	0.058	0.082	0.061	0.085	0.065
전남	0.062	0.059	0.076	0.059	0.050	0.061	0.049	0.068	0.073	0.063	0.060	0.078	0.067
경북	0.063	0.065	0.077	0.065	0.063	0.062	0.056	0.061	0.071	0.068	0.053	0.054	0.055
경남	0.055	0.073	0.060	0.055	0.076	0.073	0.050	0.054	0.053	0.054	0.052	0.053	0.072
제주	0.091	0.082	0.103	0.076	0.097	0.084	0.127	0.086	0.073	0.073	0.096	0.102	0.123

〈표 3-41〉 〈가계〉 소비지출 평균의 층 및 월별 상대표준오차

	중앙값	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
전국	0.017	0.018	0.019	0.017	0.017	0.016	0.019	0.017	0.018	0.016	0.016	0.016	0.018
동부	0.019	0.020	0.022	0.018	0.019	0.017	0.021	0.018	0.020	0.018	0.018	0.018	0.019
읍면부	0.036	0.039	0.038	0.035	0.033	0.033	0.037	0.037	0.039	0.035	0.031	0.030	0.039
서울	0.042	0.042	0.046	0.042	0.041	0.038	0.042	0.042	0.041	0.042	0.038	0.039	0.043
부산	0.047	0.051	0.055	0.047	0.050	0.043	0.045	0.041	0.053	0.046	0.041	0.044	0.049
대구	0.053	0.053	0.059	0.053	0.051	0.051	0.055	0.059	0.058	0.053	0.050	0.049	0.048
인천	0.054	0.053	0.064	0.064	0.043	0.045	0.070	0.045	0.058	0.054	0.038	0.041	0.065
광주	0.055	0.055	0.076	0.053	0.088	0.047	0.055	0.052	0.061	0.048	0.047	0.060	0.059
대전	0.058	0.057	0.075	0.056	0.064	0.055	0.058	0.077	0.058	0.054	0.067	0.054	0.060
울산	0.072	0.085	0.087	0.066	0.086	0.097	0.087	0.063	0.070	0.051	0.051	0.073	0.066
경기/동부	0.050	0.055	0.058	0.046	0.049	0.046	0.061	0.046	0.056	0.048	0.048	0.051	0.050
경기/읍면	0.064	0.060	0.070	0.073	0.064	0.057	0.089	0.074	0.082	0.060	0.049	0.049	0.064
강원/동부	0.061	0.056	0.072	0.075	0.060	0.055	0.105	0.062	0.063	0.045	0.057	0.055	0.075
강원/읍면	0.073	0.103	0.137	0.078	0.063	0.070	0.065	0.172	0.067	0.062	0.076	0.078	0.057
충북/동부	0.083	0.089	0.111	0.103	0.077	0.080	0.114	0.134	0.086	0.080	0.068	0.075	0.076
충북/읍면	0.083	0.135	0.095	0.122	0.081	0.074	0.079	0.073	0.071	0.065	0.085	0.088	0.130
충남/동부	0.087	0.084	0.088	0.087	0.095	0.092	0.076	0.090	0.090	0.103	0.081	0.083	0.068
충남/읍면	0.104	0.156	0.089	0.090	0.105	0.084	0.092	0.108	0.111	0.104	0.125	0.084	0.171
전북/동부	0.074	0.061	0.113	0.072	0.120	0.073	0.076	0.088	0.066	0.057	0.067	0.091	0.077
전북/읍면	0.146	0.160	0.145	0.149	0.146	0.149	0.146	0.139	0.146	0.247	0.136	0.187	0.130
전남/동부	0.075	0.077	0.075	0.066	0.062	0.075	0.071	0.087	0.086	0.092	0.070	0.071	0.082
전남/읍면	0.100	0.088	0.137	0.099	0.080	0.101	0.067	0.111	0.117	0.082	0.099	0.139	0.107
경북/동부	0.074	0.086	0.084	0.083	0.081	0.068	0.071	0.067	0.084	0.062	0.065	0.073	0.075
경북/읍면	0.104	0.099	0.133	0.105	0.104	0.112	0.091	0.108	0.119	0.131	0.088	0.082	0.082
경남/동부	0.075	0.098	0.079	0.072	0.104	0.095	0.064	0.072	0.068	0.076	0.071	0.074	0.099
경남/읍면	0.081	0.098	0.089	0.080	0.089	0.105	0.081	0.074	0.082	0.049	0.060	0.059	0.079
제주/동부	0.104	0.099	0.090	0.095	0.119	0.104	0.157	0.104	0.088	0.079	0.117	0.124	0.153
제주/읍면	0.109	0.123	0.312	0.093	0.116	0.110	0.112	0.108	0.102	0.171	0.088	0.094	0.074

(3) 목표 CV별 표본 크기

향후 경제활동인구조사 표본설계에서 목표정도에 따른 적정 표본크기를 산출해 보기 위해 여기서는 연동 교체가 없는 4, 8, 12월에 응답한 가구수의 평균을 현재의 표본 가구 규모로 보고, 앞 절에서 구한 상대표준오차의 중앙값을 현재의 상대표준오차(CV)로 간주하여 시도별로 목표 CV에 따른 표본크기를 다음과 같은 방식으로 산출한 결과는 <표 3-42>와 같다.

$$n^* = n_{old} \times \left(\frac{cv_{old}}{cv^*} \right)^2$$

여기서 경제활동인구조사의 전국 표본 가구수는 시도별 가구수의 합이 아닌 전국 CV를 기준으로 산정한 값이며, 아래 표에서 음영으로 표시한 부분은 현재 수준에 가까운 표본크기이다.

<표 3-42> <경향> 목표 CV별 표본 가구수

	현재 표본가구 (n)	목표 CV								
		0.01	0.02	0.03	0.04	0.05	0.08	0.10	0.12	0.15
전국	30,102	317,310	79,328	35,257	19,832	12,692	4,958	3,173	2,204	1,410
서울	3,770	233,594	58,399	25,955	14,600	9,344	3,650	2,336	1,622	1,038
부산	2,016	289,486	72,371	32,165	18,093	11,579	4,523	2,895	2,010	1,287
대구	1,653	258,723	64,681	28,747	16,170	10,349	4,043	2,587	1,797	1,150
인천	1,756	169,346	42,336	18,816	10,584	6,774	2,646	1,693	1,176	753
광주	1,399	305,184	76,296	33,909	19,074	12,207	4,769	3,052	2,119	1,356
대전	1,496	279,304	69,826	31,034	17,456	11,172	4,364	2,793	1,940	1,241
울산	1,364	280,150	70,037	31,128	17,509	11,206	4,377	2,801	1,945	1,245
경기	3,611	233,337	58,334	25,926	14,584	9,333	3,646	2,333	1,620	1,037
강원	1,696	516,177	129,044	57,353	32,261	20,647	8,065	5,162	3,585	2,294
충북	1,694	526,876	131,719	58,542	32,930	21,075	8,232	5,269	3,659	2,342
충남	1,660	389,483	97,371	43,276	24,343	15,579	6,086	3,895	2,705	1,731
전북	1,526	506,555	126,639	56,284	31,660	20,262	7,915	5,066	3,518	2,251
전남	1,707	460,805	115,201	51,201	28,800	18,432	7,200	4,608	3,200	2,048
경북	2,070	465,955	116,489	51,773	29,122	18,638	7,281	4,660	3,236	2,071
경남	1,802	391,416	97,854	43,491	24,463	15,657	6,116	3,914	2,718	1,740
제주	882	495,188	123,797	55,021	30,949	19,808	7,737	4,952	3,439	2,201

향후 표본설계를 위해 목표 정도는 전국 3%, 서울, 경기 8%, 광역시 12%, 기타도 15%로, 현재와 비슷하거나 조금 낮게 설정하였고, 집락크기가 20가구일 때와 2-1절에서 분석한 결과를 이용하여 집락크기를 15가구로 줄였을 경우 필요한 적정 가구수와 조사구수를 함께 산출한 결과는 <표 3-43>과 같다.

집락크기가 15가구일 때의 표본크기 산정을 위해 2-2절에서 구한 R_a^2 를 이용하여 $deff$ 를 구하였고, 다음 식을 이용하여 SRS 표본 대비 조사구내 가구가 20가구일 때와 동일한 효율을 갖도록 계산하였다. 조사구수는 가구수를 각각의 집락 크기로 나눈 값이고, 연동그룹의 수에 따른 조정 등은 반영하지 않은 결과이다.

$$deff_{(M)} = \frac{V(\widehat{t_{cluster}})}{V(\widehat{t_{SRS}})} \approx [1 + (M-1)ICC], \quad ICC \approx R_a^2$$

$$n_{SRS} = n_{(20)} / deff_{(20)}, \quad n_{(15)} = n_{SRS} \times deff_{(15)}$$

<표 3-43> <경찰> 시도별 적정 표본가구수와 조사구수 1

	현재		CV		집락 크기(20)		집락 크기(15)	
	가구수	조사구수	현재	목표	가구수	조사구수	가구수	조사구수
전국	30,102	1,629	0.032	0.03	34,961	1,748	33,255	2,217
서울	3,770	207	0.079	0.08	3,650	182	3,359	224
부산	2,016	108	0.120	0.12	2,010	101	1,925	128
대구	1,653	90	0.125	0.12	1,797	90	1,714	114
인천	1,756	99	0.098	0.12	1,176	59	1,122	75
광주	1,399	81	0.148	0.12	2,119	106	1,983	132
대전	1,496	81	0.137	0.12	1,940	97	1,813	121
울산	1,364	72	0.143	0.12	1,945	97	1,865	124
경기	3,611	198	0.080	0.08	3,646	182	3,560	237
강원	1,696	90	0.174	0.15	2,294	115	2,164	144
충북	1,694	90	0.176	0.15	2,342	117	2,246	150
충남	1,660	90	0.153	0.15	1,731	87	1,659	111
전북	1,526	81	0.182	0.15	2,251	113	2,159	144
전남	1,707	90	0.164	0.15	2,048	102	1,986	132
경북	2,070	108	0.150	0.15	2,071	104	1,933	129
경남	1,802	99	0.147	0.15	1,740	87	1,647	110
제주	882	45	0.237	0.15	2,201	110	2,121	141

한편, <표 3-44>는 목표 CV에 따른 가계동향조사에서의 표본 가구수를 정리한 것이다. 가계동향조사의 전국 표본 가구수도 시도별 가구수의 합이 아닌 전국 CV를 기준으로 산정한 값이며, 아래 표에서 음영으로 표시한 부분은 현재 수준에 가까운 표본크기이다.

<표 3-44> <가계> 목표 CV별 표본가구수

	현재 표본가구 (n)	목표 CV						
		0.01	0.02	0.03	0.04	0.05	0.08	0.10
전국	7,064	20,151	5,038	2,239	1,259	806	315	202
동부	5,588	19,613	4,903	2,179	1,226	785	306	196
읍면부	1,475	18,984	4,746	2,109	1,187	759	297	190
서울	852	14,954	3,738	1,662	935	598	234	150
부산	493	10,780	2,695	1,198	674	431	168	108
대구	414	11,515	2,879	1,279	720	461	180	115
인천	389	11,179	2,795	1,242	699	447	175	112
광주	375	11,182	2,795	1,242	699	447	175	112
대전	394	13,242	3,310	1,471	828	530	207	132
울산	257	13,200	3,300	1,467	825	528	206	132
경기	780	14,462	3,615	1,607	904	578	226	145
강원	460	11,075	2,769	1,231	692	443	173	111
충북	368	15,552	3,888	1,728	972	622	243	156
충남	452	22,124	5,531	2,458	1,383	885	346	221
전북	375	16,533	4,133	1,837	1,033	661	258	165
전남	384	14,658	3,665	1,629	916	586	229	147
경북	452	17,741	4,435	1,971	1,109	710	277	177
경남	440	13,191	3,298	1,466	824	528	206	132
제주	179	14,813	3,703	1,646	926	593	231	148

한편 경제활동인구조사에서 조사구당 15가구(3개 구역)를 조사한다고 가정하고, 현재와 같이 경제활동인구조사의 표본 조사구에서 10가구(2개 구역)를 대상으로 가계동향조사를 실시한다고 가정하면, 가계동향조사에서 시도별 표본크기와 예상 CV는 <표 3-45>와 같다. 제시된 표본 가구수는 조사구수에 조사구당 표본 가구수인 10을 곱한 값이며, 이 경우 약 22,000가구가 조사대상이 되고 이때 CV는 전국 1%, 동부 1.2%, 동부 1.5% 수준으로 예상할 수 있다.

〈표 3-45〉 〈가계〉 시도별 적정 표본가구수와 조사구수 1

	현재		CV		집락 크기(10)	
	가구수	조사구수	현재	예상	가구수	조사구수
전국	7,064	999	0.017	0.010	22,170	2,217
동부	5,588	783	0.042	0.012	13,920	1,392
읍면부	1,475	216	0.047	0.015	8,250	825
서울	852	126	0.053	0.026	2,240	224
부산	493	72	0.054	0.029	1,280	128
대구	414	54	0.055	0.032	1,140	114
인천	389	54	0.058	0.039	750	75
광주	375	54	0.072	0.029	1,320	132
대전	394	54	0.043	0.033	1,210	121
울산	257	36	0.049	0.033	1,240	124
경기	780	108	0.065	0.025	2,370	237
강원	460	63	0.070	0.028	1,440	144
충북	368	54	0.066	0.032	1,500	150
충남	452	63	0.062	0.045	1,110	111
전북	375	54	0.063	0.034	1,440	144
전남	384	54	0.055	0.033	1,320	132
경북	452	63	0.091	0.037	1,290	129
경남	440	63	0.052	0.035	1,100	110
제주	179	27	0.083	0.032	1,410	141

한편, 가계동향조사는 경제활동인구조사의 조사구에서 10가구를 조사하도록 설계하였지만, 조사구가 부적격인 경우 제외하므로, 현행 조사에서 경제활동인구조사의 조사대상인 1,629개 조사구 중 999개(61%)의 조사구가 가계동향조사의 조사대상임을 감안하여, 2,217개 조사구의 약 61%에 해당하는 1,359개 조사구를 적격 조사구로 가정하고 가계동향조사에서 표본크기와 예상 CV를 산출한 결과는 <표 3-46>과 같다.

〈표 3-46〉 〈가계〉 시도별 적정 표본가구수와 조사구수 2

	현재		CV		집락 크기(10)	
	가구수	조사구수	현재	예상	가구수	조사구수
전국	7,064	999	0.014	0.012	13,590	1,359
동부	5,588	783	0.015	0.015	8,530	853
읍면부	1,475	216	0.033	0.019	5,060	506
서울	852	126	0.037	0.033	1,370	137
부산	493	72	0.045	0.037	780	78
대구	414	54	0.049	0.041	700	70
인천	389	54	0.051	0.049	460	46
광주	375	54	0.051	0.037	810	81
대전	394	54	0.052	0.042	740	74
울산	257	36	0.067	0.042	760	76
경기	780	108	0.038	0.031	1,460	146
강원	460	63	0.048	0.035	880	88
충북	368	54	0.058	0.041	920	92
충남	452	63	0.057	0.057	680	68
전북	375	54	0.061	0.043	880	88
전남	384	54	0.059	0.043	810	81
경북	452	63	0.054	0.047	790	79
경남	440	63	0.052	0.044	670	67
제주	179	27	0.083	0.041	870	87

* 적격가구: 농가(임가포함) 및 어가, 음식점, 여관, 하숙업을 경영하는 겸용주택내 가구, 가구원 중 영업상 사용인이 2인 이상 동거하는 가구, 비혈연 가구원으로 구성된 가구, 외국인 가구 등을 제외한 가구

* 각 도의 동부, 읍면부별로 조사구당 적격가구수가 동부는 평균 57.1가구인 반면, 읍면부는 29.8가구로 차이가 있으며, 기존 설계에서는 조사구당 예상 적격가구수가 4가구 미만인 조사구는 제외하였음

3) 연동 모형 개선을 위한 제언

(1) 연동 교체 방안

현재의 경제활동인구조사 연동표본은 전체 표본을 9개 그룹으로 구분하고 조사구 내 표본 가구를 4구역으로 나누어, 매월 181개 조사구(905가구)를 교체하는 방식으로 표본이 교체된다. 따라서 현행 연동 모형에 따르면 모든 표본 가구는 36개월 동안 조사에 참여한 후 표본에서 제외되고 이를 대체하는 새로운 가구(조사구)가 표본에 유입되는 연동 방식이 적용되고 있다.

현재 연동 방식은 우선 가구부문 연동 표본의 모집단이 되는 인구주택총조사의 조사주기가 5년이라는 점을 고려할 때 교체 주기가 36개월을 2회 유지할 수 없기 때문에 인구주택총조사 주기와 연동주기가 일치하지 않는다는 한계를 갖고 있다. 아울러 한 조사구(4개 구역)가 대체되는데 4개월이 소요 되고 특히 가계동향조사 표본의 경우 조사구 교체 기간 중 3번째 및 4번째 달에 표본 가구(구역)를 교체하기 때문에 가구동향조사의 분기별 통계 작성 주기와 조사구 교체 주기가 어긋난다는 문제도 갖고 있다.

따라서 본 연구에서는 앞에서 분석한 집락효과 분석 내용을 토대로 집락당(조사구당) 표본 가구수를 15가구로 조정하고, 전체 표본을 10개 연동 그룹으로 구분하여, 표본 가구를 30개월 동안 조사한 후에 교체하는 방안을 검토하기로 한다. 여기서 집락의 크기를 15가구로 가정하고 30개월 단위로 표본을 연동함에 따라 여기서는 <표 3-43>에 제시된 목표 정도에 따른 구한 표본 조사구수를 10의 배수가 되도록 조정하는 것을 가정하고 연동 방안을 검토하기로 한다. 총별 표본 조사구수를 10의 배수로 조정하는 새로운 연동 모형에서의 표본크기는 <표 3-47>과 같으며, 이 경우 표본크기가 기존의 표본설계에 비해 증가하게 된다는 점을 참고하기 바란다.

〈표 3-47〉 〈경찰〉 시도별 적정 표본가구수와 조사구수 2

	현재		CV		집락 크기(15)		10의 배수로 조정	
	가구수	조사구수	현재	목표	가구수	조사구수	가구수	조사구수
전국	30,102	1,629	0.032	0.03	33,255	2,217	35,100	2,340
서울	3,770	207	0.079	0.08	3,359	224	3,450	230
부산	2,016	108	0.120	0.12	1,925	128	1,950	130
대구	1,653	90	0.125	0.12	1,714	114	1,800	120
인천	1,756	99	0.098	0.12	1,122	75	1,200	80
광주	1,399	81	0.148	0.12	1,983	132	2,100	140
대전	1,496	81	0.137	0.12	1,813	121	1,950	130
울산	1,364	72	0.143	0.12	1,865	124	1,950	130
경기	3,611	198	0.080	0.08	3,560	237	3,600	240
강원	1,696	90	0.174	0.15	2,164	144	2,250	150
충북	1,694	90	0.176	0.15	2,246	150	2,400	160
충남	1,660	90	0.153	0.15	1,659	111	1,800	120
전북	1,526	81	0.182	0.15	2,159	144	2,400	160
전남	1,707	90	0.164	0.15	1,986	132	2,100	140
경북	2,070	108	0.150	0.15	1,933	129	2,100	140
경남	1,802	99	0.147	0.15	1,647	110	1,800	120
제주	882	45	0.237	0.15	2,121	141	2,250	150

한편 새로운 30개월 기준 연동 모형을 도입하면 기존의 36개월 기준 연동 표본과 새로운 연동 모형을 연계하기 위해 표본의 유지 기간이 다르다는 점을 고려해야 한다. 구표본이 현행과 같이 36개월 동안 동일한 규모로 이탈하게 되면 새로운 연동표본이 모두 진입한 후에도 구표본이 남아 있게 된다. 따라서 구표본의 조사대상 가구가 표본으로 유지되는 기간이 감소하더라도 30개월 동안에 구표본과 신표본이 동시에 완전하게 교체되는 방안을 구현하는 것이 필요하다.

이런 측면을 고려해 구표본의 이탈 시기에 따라 3가지 연계 방안을 검토해 보기로 한다. 표본 관리의 편의를 위해 같은 그룹 내에서는 구역별로 조사기간이 동일하게 유지되도록 하였고, 신표본은 30개월 동안 매월 234개 조사구(1,170가구)가 진입하도록 설계하였다.

[1안]은 최초 11개월까지는 현재의 방법을 유지하며 교체가 완료되는 30번째 달을 기준으로 역산하여 구표본을 그룹별로 1개월씩 빠르게 이탈시키는 방법이다. <표 3-48>과 같이 1-3그룹은 36개월, 4그룹은 35개월, 5그룹은 34개월, 6그룹은 33개월, 7그룹은 32개월, 8그룹은 31개월, 9그룹은 30개월 조사 후에 이탈하도록 하면, 30개월 동안 구표본과 신표본의 교체가 이루어진다. 이때, 전체 표본 가구수는 최소

32,845가구에서 최대 35,495가구로 구표본의 32,580가구 보다 규모가 크기 때문에 안정적이지만, 신표본의 목표 가구수인 35,100가구보다 큰 시기에는 조사비용이 과다 지출될 수 있다는 약점이 있다.

[2안]은 신표본이 들어오는 첫 번째 달부터 구표본의 유지 기간을 줄여나가는 방법이다. <표 3-49>와 같이 1그룹은 35개월, 2그룹은 34개월, 3그룹은 33개월, 4그룹은 32개월, 5그룹은 31개월, 6-9그룹은 30개월 조사 후에 이탈하게 된다. 이때, 전체 표본 가구수는 최소 31,125가구에서 최대 35,100가구로 처음 20개월 동안은 구표본의 32,580가구 보다 규모가 작지만, 교체가 진행되면서 점점 증가하기 때문에 21개월부터는 구표본의 표본 규모 이상을 유지할 수 있다.

[3안]은 1안과 2안을 절충한 방법으로 네 번째 달부터 구표본의 유지기간을 줄여나가는 방법이다. <표 3-50>과 같이 1그룹은 36개월, 2그룹은 35개월, 3-4그룹은 34개월, 5그룹은 33개월, 6그룹은 32개월, 7그룹은 31개월, 8-9그룹은 30개월 조사 후에 이탈하게 된다. 전체 표본 가구수는 최소 32,625가구에서 최대 35,100가구로써 구표본의 32,580가구 보다 항상 크게 유지된다는 장점이 있으며, 2안 보다는 안정적이고, 1안 보다는 조사비용이 절감된다는 특징이 있다. 만약 새로운 30개월 기준 연동모형을 도입한다면 구표본과의 연계를 위해서는 [3안]이 가장 현실적인 연동 모형 연계 방안이 될 수 있을 것으로 생각된다.

물론 이런 새로운 30개월 연동모형을 도입하기 위해서는 기존 36개월 기준 연동모형에 비해 업무량이 늘어난다는 문제가 발생하기 때문에 정부차원에서 관련 인력이나 예산을 추가로 지출할 의지를 갖는 경우에만 실현 가능하다는 점을 다시 한번 강조하고자 한다. 현재 보다 정확한 노동력 관련 통계에 생산에 대한 요구가 늘어나고 있을 뿐만 아니라 보다 세분화된 노동력 통계에 대한 수요가 급증하고 있다는 점을 고려할 때, 본 연구에서 제시된 연동모형 도입 방안을 검토해 보는 것도 큰 의미가 있을 것으로 보인다.

〈표 3-48〉 연동표본 교체 방안(1안)

[illegible]

*구표본의 조사기간 변경(1-3그룹: 36개월, 4그룹 : 35개월, 5그룹 : 34개월, 6그룹 : 33개월, 7그룹 : 32개월, 8그룹 : 31개월, 9그룹 : 30개월)

〈표 3-49〉 연동표본 교체 방안(2안)

[illegible]

*구표본의 조사기간 변경(1그룹 : 35개월, 2그룹 : 34개월, 3그룹 : 33개월, 4그룹 : 32개월, 5그룹 : 31개월, 6-9그룹 : 30개월)

〈표 3-50〉 연동표본 교체 방안(3안)

[illegible]

*구표본의 조사기간 변경(1그룹 : 36개월, 2그룹 : 35개월, 3-4그룹 : 34개월, 5그룹 : 33개월, 6그룹 : 32개월, 7그룹 : 31개월, 8-9그룹 : 30개월)

IV. 추정량 및 분산추정량 산출방안 연구

본 장에서는 각 연동방안으로부터 수집된 자료를 이용하여 조사 시점별(월별, 분기별 등) 모수에 대한 추정량과 추정량의 분산을 산출하는 방안들에 대하여 살펴보도록 한다. 1절에서는 먼저 고려할 수 있는 가능한 추정량과 이들의 성질들을 살펴보자. (Philip Bell, Comparison of Alternative Labour Force Survey Estimators), (WA Fuller, A regression composite estimator with application to the Canadian Labour Force Survey), (Philip A. Bell and Anthony M. Carolan, TREND ESTIMATION FOR SMALL AREAS from a continuing survey with controlled sample overlap). 이어 그림에서는 현행 1/36 연동모형하에서 사용되고 있는 사후총화추정량의 연동그룹별 시계열 특성을 파악하여 편향추정방안의 통계적 타당성을 평가하고자 한다. 3절에서는 현재 사용하고 있는 분산추정량의 통계적 타당성을 이론적으로 평가하고 또한 실제 경제활동인구조사자료 및 모의실험을 통해 적절한 분산추정량을 제시하였다. 마지막 4절에서는 사후총화추정량 대신 고려할 수 있는 방안으로 시계열모형에 근거한 최량선형불편추정량을 고려하고 이를 위해 3년간의 월별 실업자를 시계열 자료로 분석하고 이의 사용타당성을 분석하였다.

1. 추정량의 종류

(1) 불편추정량

- 각 관측치에 추출확률을 고려하여 모집단을 추정한 방법으로 t 시점의 모집단 총합에 대한 추정량은 다음과 같음.

$$\hat{y}_t^H = \sum_i w_{ti}^{\pi} y_{ti}$$

- w_{ti}^{π} 표본가중치 (sampling weight)를 의미함.
- 연동표본을 고려하지 않음.
- 연동표본 특성을 고려한 연동그룹 추정량은 다음과 같음.

$$\begin{aligned} \hat{y}_t^{Hr} &= \frac{1}{R} \sum_{i: R(t,i)=r}^R c w_{ti}^{\pi} y_{ti}, \quad i = 1, 2, \dots, R \\ &= \frac{1}{R} \sum_{i: R(t,i)=r}^R \hat{y}_t^{Rr} \end{aligned}$$

- * 각 연동그룹 추정량 $\hat{y}_t^{Rr}$ 을 이용한 방법.
- * $R(t,i)$ 는 t 시점에서 i 번째 연동그룹을 의미함.
- * 각 연동그룹은 서로 다른 응답횟수를 가진 표본들로 구성됨.
- * 상수 c 는 연동표본 내 추출확률을 보정해주는 상수. 가령, 1개의 연동그룹 안에 560개의 층을 모두 고려하게 되면, 모집단 수에 비해 연동그룹 표본수가 작기 때문에 표본 내 층의 수는 약 71개 정도로 나눠짐. 이 때 추출확률에 상수 $c(\frac{560}{71} \approx 8)$ 를 곱해서 연동그룹 추정량을 구함.

(2) 사후층화(회귀) 추정량

- 일반화 회귀 추정량(Generalised Regression Estimator : GR)
 - 관심변수 Y 의 추정량에 지역, 성별, 연령 등과 같은 모집단 보조정보 X 를 이용하여 보정하는 방법
 - 모집단 보조정보는 표본 밖으로부터 온 자료이므로 추정량의 질을 개선 시킴.

$$\hat{y}_t^G = \hat{y}_t + (X_t - \hat{x}_t)\hat{\beta},$$

여기서

$$\hat{\beta}_t = \left(\sum_i b_{ti} x_{ti}' x_{ti} \right)^{-1} \sum_i b_{ti} x_{ti}' y_{ti},$$

$$\hat{x}_t = \sum_i b_{ti} x_{ti}, \quad \hat{y}_t = \sum_i b_{ti} y_{ti},$$

- x_t 는 보조변수들의 모집단 총합을 나타내며 $(\hat{x}_t, \hat{y}_t)$ 은 시점 t 의 불편추정량을 나타낸다. 회귀계수 추정에 사용되는 b_{ti} 의 값으로는 가능한 여러 가지 방안들이 있으며 흔히 표본가중치가 사용됨.
- 모집단 보조정보가 각 관측치의 사후층 포함여부를 나타내는 지시함수들의 벡터인 경우 사후층화 추정방법과 동일함.
- 관심변수 Y 의 모집단 총합에 대한 사후층화 추정량 (일본: 4-8-4)

$$\hat{y}_t^{post} = \hat{y}_t + (X_t - \hat{x}_t)\hat{\beta}$$

$$\hat{\beta}_t = \left(\sum_i w_{ti}^{\pi} x_{ti}' x_{ti} \right)^{-1} \sum_i w_{ti}^{\pi} x_{ti}' y_{ti}$$

- * 각 월별 통계량의 통계적 모형화를 배제한 당월 자료와 모집단 정보를 이용한 추정량으로 쉽게 적용(easy to implement) 할 수 있는 추정량.
- * 일본의 경우, 각 관측치의 성, 연령 그리고 거주 지역을 보조정보로 사용함.

(3) AK 복합추정량(Composite estimator)

- 이전 월 자료와 금월의 자료를 결합하여 계산하는 복합추정방법 (4-8-4 모형)
- AK 복합추정방법은 연동그룹편향이 있는 경우, 경험적 결과를 바탕으로 편향을 보정하는 효과가 있음. K 를 이용하여 지속표본(matched rotation group) 내에서의 변화량을 고려하고, A 값을 이용하여 진입표본과 계속표본간의 차이를 보정함.
- 적절한 A 값과 K 값의 선택이 중요함.
- 4-8-4 교체방안 하에서의 추정량은 다음과 같음.

$$\hat{y}_t^{AK} = (1-K) \frac{1}{8} \sum_{r=1}^8 \hat{y}_t^{Rr} + K(Y_{t-1}^{AK} + \Delta_t) + A\hat{\beta}_t$$

$\hat{y}_t^{Rr}$: t 시점에서의 r 번째 연동그룹 추정량

$$\Delta_t = \frac{8}{6} \sum_{r \in S} (\hat{y}_t^{Rr} - \hat{y}_{t-1}^{Rr}) : \text{지속표본의 변동분}$$

$$\hat{\beta}_t = \sum_{i \notin S} \hat{y}_t^{Rr} - \frac{2}{6} \sum_{i \in S} \hat{y}_t^{Rr} : \text{신규표본 지속표본 그룹간의 차이}$$

$S = \{2, 3, 4, 5, 6, 7, 8\}$: 전월에 이어 연속 조사하는 그룹

- * Δ_t 는 지속표본그룹의 변동 분으로, 이전월과 금월의 지속표본그룹은 6개의 그룹으로 구성. 추정량 Y_{t-1}^{AK} 는 8개의 그룹의 합이므로, 변동분에 $\frac{8}{6}$ 을 곱해주어 두 값의 크기를 조정.

- * $\hat{\beta}_t$ 는 신규표본과 지속표본 그룹 간 차이를 의미. 2개의 신규표본과 6개의 지속표본들의 차이를 추정하기 위해, 6개의 지속표본들의 합에 $\frac{2}{6}$ 을 곱해서 크기를 조정.

- $A=0, K=0.5$ 인 AK 복합추정량 (미국 : 4-8-4 교체방안)

$$\hat{y}_t^{AK} = (1-K) \frac{1}{8} \sum_{r=1}^8 \hat{y}_t^{Rr} + K(Y_{t-1}^{AK} + \Delta_t)$$

$\hat{Y}_t$: 해당 월의 추정량

Δ_t : 해당 월(t)과 이전 월(t-1) 지속표본그룹의 변화량

• 단점

- 가중값 A와 K선택에 어려움이 있음.
- AK추정량에 있어서 가장 중요한 모수는 K이다. 즉, 계속표본들의 변화를 얼마나 반영하는지를 나타내는 모수 값을 의미한다. Optimal한 A와 K는 정해져 있는 것이 아니라 어떤 변수를 추정할 것인지에 따라 달라진다. 미국의 경우 취업률을 추정하는데 있어서는 K값이 1에 가까울수록 좋은 것으로 알려져 있으며 이는 취업자들이 실업자가 되는 경우보다는 계속 취업인 상태가 많기 때문에 서로 같은 연동 그룹 내에서의 취업률은 상관관계가 매우 높기 때문이다. 이와 같이 상태변화가 없는 변수에 대해서는 현재 시점에 수집된 응답보다는 과거자료에 계속표본들의 변화량을 고려한 두 번째 추정량에 더 큰 가중 값을 사용하는 것이 더 적절할 것으로 판단된다.
- AK추정량은 현시점 추정량 산출에서 이전시점의 자료들도 사용된다. 만일, 과거에 조사되지 않은 새로운 변수에 대한 추정이 필요한 경우, 이

전시점들에서의 AK복합추정량을 모두 산출해야하기 때문에 매우 복잡하고 불편하다.

(4) A Modified 회귀추정량

- 기존 GR추정량의 변형방법으로 제어변수(Control variable)를 이용하는 특징을 가짐. 제어변수(z)와 모집단 보조정보(x)를 모두 고려하는 일반화 회귀추정방법.
- 제어변수(Z)를 사용하여 조사 자료에서 y값을 먼저 보정하는 방법으로, 이 방법의 핵심은 조사 자료 내에서 관심변수(Y)를 잘 설명할 수 있는 적절한 기준변수(key item; y_t^*)를 사용하는 것임.
- 1/6 교체방안 하에서의 추정량은 다음과 같음.

$$\hat{y}_t^M = \hat{y}_t^H + ((X_t, Z_t) - (\hat{x}_t^H, \hat{z}_t^H))\beta_t^M$$

$$\beta_t^M = \left(\sum_i w_{ti}^\pi (x_{ti}, z_{ti})' (x_{ti}, z_{ti}) \right)^{-1} \sum_i w_{ti}^\pi (x_{ti}, z_{ti})' y_{ti}$$

- 제어변수는 다음과 같이 정의됨.

$$z_{ti} = \begin{cases} (1-a)\frac{6}{5}y_{t-1,i}^* + a\left(y_{ti}^* - \frac{6}{5}(y_{ti}^* - y_{t-1,i}^*)\right) & \text{for } i \in D \\ ay_{t,i}^* & \text{for } i \notin D \end{cases}$$

- $\frac{5}{6}$ 은 중복되는 표본의 비율을 나타내고, $D = \{2, 3, 4, 5, 6\}$ 으로 지속표본을 의미함.
- 여기서 $\hat{x}_t^H, \hat{y}_t^H$ 는 w_{ti}^π 표본가중치 (sampling weight)를 이용한 불편추정량.
- $\hat{z}_t^H = (1-a)\frac{6}{5}\hat{y}_{t-1}^{*HD} + a(y_t^{*HD} - \frac{6}{5}(y_t^{*HD} - y_{t-1}^{*HD}))$ 으로 추정량들의 선형결합
- $\hat{y}_t^{*H}$ 는 기준 변수(key item)을 y_t^{*HD}, y_{t-1}^{*HD} 는 지속표본그룹으로 계산된 당월, 전월의 추정량을 의미함.
- 제어변수(z)는 a=0인 경우 MR1추정방법을, a=1인 경우 MR2추정방법을 의미.
- 그밖에 $a \in [0, 1]$ 인 a를 이용하여 제어변수를 조정하는 MF 추정방법이 있음.

- MR1 추정방법

$$\hat{y}_t^M = \hat{y}_t^H + ((X_t, Z_t) - (\hat{x}_t^H, \hat{z}_t^H))\beta_t^M$$

$$\beta_t^M = \left(\sum_i w_{ti}^\pi(x_{ti}, z_{ti})' (x_{ti}, z_{ti}) \right)^{-1} \sum_i w_{ti}^\pi(x_{ti}, z_{ti})' y_{ti}$$

- 제어변수는 다음과 같이 정의됨.

$$z_{ti} = \begin{cases} \frac{6}{5} y_{t-1,i}^* & \text{for } i \in D \\ 0 & \text{for } i \notin D \end{cases}$$

$$\hat{z}_t^H = \begin{cases} \frac{6}{5} \hat{y}_{t-1}^{*HD} & \text{for } i \in D \\ 0 & \text{for } i \notin D \end{cases}$$

- MR2 추정방법 (캐나다 : 1/6교체 방안)

$$\hat{y}_t^M = \hat{y}_t^H + ((X_t, Z_t) - (\hat{x}_t^H, \hat{z}_t^H))\beta_t^M,$$

$$\beta_t^M = \left(\sum_i w_{ti}^\pi(x_{ti}, z_{ti})' (x_{ti}, z_{ti}) \right)^{-1} \sum_i w_{ti}^\pi(x_{ti}, z_{ti})' y_{ti}$$

- 제어변수는 다음과 같이 정의됨.

$$z_{ti} = \begin{cases} y_{ti}^* - \frac{6}{5} (y_{ti}^* - y_{t-1,i}^*) & \text{for } i \in D \\ y_{t,i}^* & \text{for } i \notin D \end{cases}$$

$$\hat{z}_t^H = \begin{cases} \hat{y}_t^H - \frac{6}{5} (\hat{y}_t^{*HD} - \hat{y}_{t-1}^{*HD}) & \text{for } i \in D \\ \hat{y}_t^H & \text{for } i \notin D \end{cases}$$

- MR추정량 사용에 있어서 다음과 같은 경우를 고려해야함.

- 제어변수는 이전시점의 기준변수(y_t^*)들의 추정량으로 이전시점 기준변수 추정량과 현재시점 기준변수 추정량에 지속표본들의 변동을 보정한 추정량들의 가중평균으로 표현된다. 이 제어변수 추정에 있어서 지속표본들의 변동만 고려하기 때문에 생기는 다음의 두 가지 문제가 있다. 첫째로, 진입표본과 지속표본이 다른 성향을 가지게 될 경우 **matched sample bias**가 발생하게 된다. 이는 제어변수가 진입표본을 고려하지 않음으로 인해서 생기는 오차이다. 두 번째로 MR2추정량에서 발생하는 문제인데 K번째 기준변수(y_t^*)가 매월 비슷한 값을 갖는 경우 제어변수(z)는 동일한 값에 수렴하게 되고, 이로 인해 모든 시점에서 관심변수 추정량은 모두 같아진다. 따라서 한번 반영된 제어변수들은 계속해서 추정량에 영향을 미치게 된다.

- AK추정량과의 비교
 - MR추정량은 진입표본을 고려하지 않고, 계속표본들의 변동만 고려한다는 점에서 AK추정량과 비슷하다. 하지만 MR추정량의 오차가 더 작고, Optimal한 A, K값을 따로 찾지 않아도 된다. 이와 같은 이유로 AK복합추정량보다 적용하기 쉽다는 장점을 가진다.

(5) 최량선형불편추정량(Best Linear Unbiased Estimator; BLUE)

- 각 연동그룹별 추정량들의 분산 공분산 행렬을 이용하여 BLUE를 구함.

$$\hat{y}_t^B = \sum_{s=t-l}^t \sum_{r=1}^8 a_{sr} \hat{y}_t^{Rr}$$

- a_{tr} 는 $var(\hat{y}_t^B)$ 을 최소화시키는 모수(parameter)이며, 당월자료와 전월 자료에 대하여 각각 다음을 만족함.

$$\sum_{r=1}^8 a_{sr} = \begin{cases} 1, & s = t \\ 0, & s \neq t \end{cases}$$

- 여기서 $var(\hat{y}_t^B)$ 값은 연동그룹추정량들의 분산 공분산행렬을 이용하여 추정될 수 있음.
- 과거 자료의 정보와 연동그룹별로 추정 값에 미치는 영향력을 조정하는 BLUE계수(a_{tr})를 이용하여 추정량 자체로 연동그룹편향을 조정할 수 있음.
- BLUE 추정량을 구하는 과정은 다음과 같음.
 - 당월 추정량은 R개의 연동그룹의 추정량의 평균이고 다음을 만족함.

$$\hat{y}_t = \frac{1}{R} \sum_{i: R(t,i)=r}^R \hat{y}_t^{Rr}, \quad E(\hat{y}_t) = Y_t$$

$$Y_t = \hat{y}_t + e_t, \quad t = 1, 2, \dots, T$$

- 연동그룹추정량은 당월 평균추정량에 오차와 편향의 합으로 나타낼 수 있고, 이때 연동편향(b_{tr})은 0으로 가정함.

$$\hat{y}_t^{Rr} = \hat{y}_t + e_{tr} + b_{tr}, \quad b_{tr} \approx 0$$

- 연속조사에서 연동그룹은 시간의 흐름에 따라 반복 측정된 시계열 자료임. 연동그룹간의 상관관계는 다음과 같이 정의할 수 있음.

$$corr(\hat{y}_t^{Rr}, \hat{y}_s^{Rr'}) = \begin{cases} \rho^{W_{|t-s|}} & \text{for } r-r' = t-s \\ \rho^{B_{|t-s|}} & \text{for } r-r' = t-s+8m, \quad m \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

- ρ_k^W 는 동일 연동그룹(within)들의 시차(lag k)에 따른 상관관계를 의미하고, ρ_k^B 서로 다른 표본에서 추출되었지만 동일한 성질을 갖는 연동그룹(between)들의 상관관계를 의미한다. 그 외에, 다른 연동그룹간의 상관관계는 0이라고 가정한다.
- 위와 같이 정의된 상관관계를 이용하여 분산을 추정할 수 있음. 추정방법으로 MME, AR(2) model, state space model이 제안되었다. (Pilip A. Bell and Anthony M. carolan ; Trend Estimation for Small Area)
- BLUE-B1 추정량 (오스트레일리아 : 1/8 교체 방안)
 - 과거 표본의 추정량을 고려해서 만들어진 각 연동그룹들의 선형결합 BLUE(Best Linear Unbiased Estimation)추정량에 모집단 분포를 이용하여 보정하는 방법.

$$\hat{y}_t^{B1} = \hat{y}_t^{BH} + (X_t - x_t^{BH})\hat{\beta}$$

$$\hat{\beta} = \left(\sum_{s=t-l}^t \sum_i w_{si}^{\pi} x_{si}^{*'} x_{si}^* \right)^{-1} \sum_{s=t-l}^t \sum_i w_{si}^{\pi} x_{si}^{*'} y_{si}^*$$

- y_{ti}^* 와 x_{ti}^* 는 각각 관심변수와 보조변수들의 BLUE 추정량으로 연동그룹 추정량(y_{ti})과 BLUE multiplier(a_{ti})의 곱으로 정의.

$$\hat{y}_t^{BH} = \sum_{s=t-l}^t \sum_i w_{si}^{\pi} y_{si}^*, \quad x_t^{BH} = \sum_{s=t-l}^t \sum_i w_{si}^{\pi} x_{si}^*$$

- 과거 자료의 기간이 길어지면 계산이 복잡해지고, 짧으면 오차가 커지는 단점이 생기므로 적절한 과거 자료 기간 선택이 필요함.

2. 현행 사후층화추정량 분석

현행 1/36 연동방안에서 고려하고 있는 추정량은 사후층화 추정량으로 표본내의 모든 가구 및 가구원들이 동일한 표본추출확률을 갖는다는 가정 하에서 각 개인의 성, 연령, 지역 정보를 이용하여 사후층화추정량을 고려하였다. 1/36 연동표본의 경우, 월간 혹은 분기간의 표본 중복 비율이 매우 높아 추정량의 시간에 따른 안정성이 확보된다. 또한 현재 전체 표본의 수가 약 32,000 가구 및 70,000여 명으로서 추정량의 통계적 정도 역시 높다. 다만 각 가구 및 가구원의 조사노출 기간이 길어집

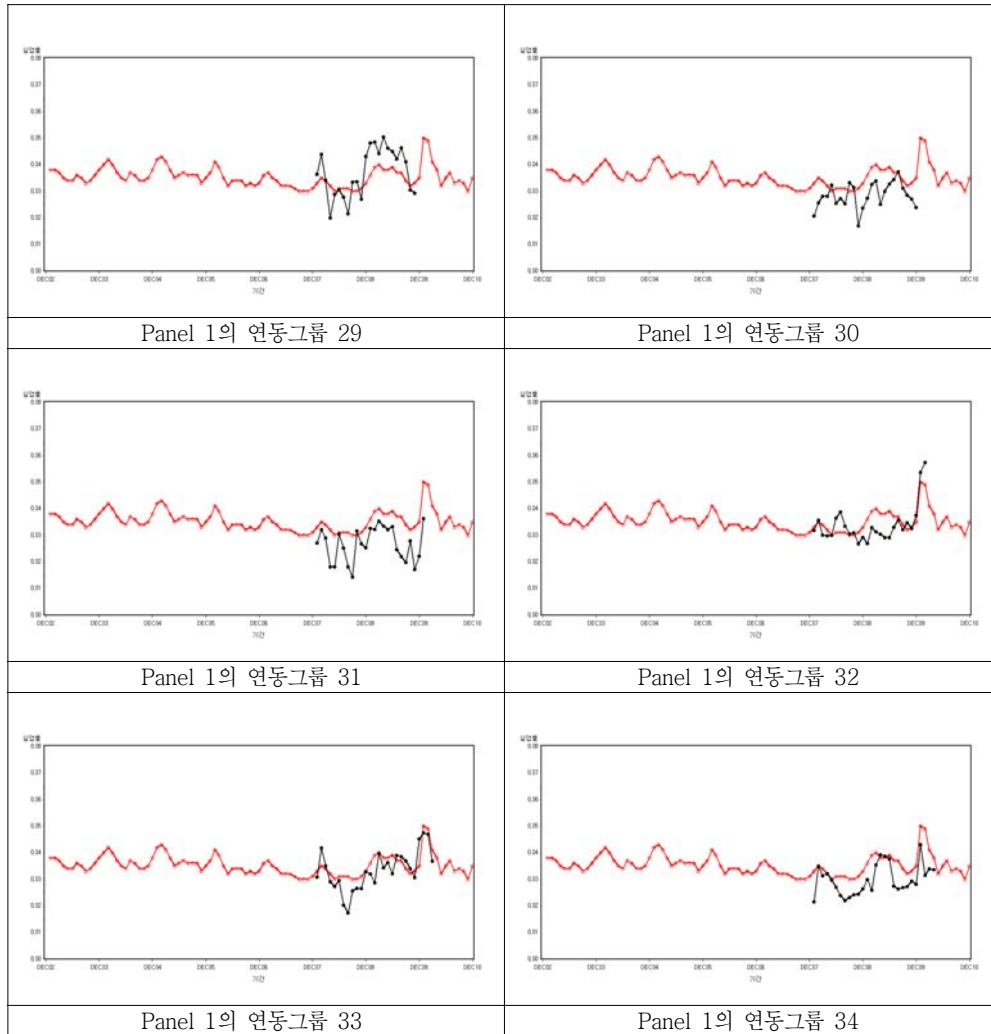
으로 인하여 일종의 패널조건화와 같은 비표본오차가 발생할 수 있을 수 있다는 문제가 제기 될 수 있다. 패널조건화로 인한 추정량의 편의 존재여부는 그 평가가 쉽지 않다. 특별히 모수의 참값을 모르는 상황에서 조사시점 관측된 모든 관측치들의 선형결합으로 추정량이 계산되어지고 따라서 각 연동그룹별 편의를 직접적으로 추정 혹은 예측하는 것은 불가능하다. 이러한 제약 하에서 본 절에서는 2008년 1월부터 2010년 12월까지의 연동그룹별 실업률 자료 분석을 통해 조사 노출 빈도에 따른 체계적 편향이 존재하는지를 살펴본다.

<표 4-1>은 72개의 연동그룹¹⁾별 2년간(2008년도 1월~2010년 12월)의 실업률 산출 결과이다. 실업률의 계산은 통계청에서 제공한 가중치를 이용하여 각 월*연동그룹별 실업인구 추정량 대 경제활동인구 추정량의 비율로 정의하였다. 동일한 연동그룹이 최대 36개월 경제활동인구조사에 노출되며 따라서 계산되어진 추정량들을 통해 일정한 시계열 구조를 갖게 되는지 예측할 수 있다. [그림 4-1]은 각 연동그룹별 3년간의 시 도표(time plot)를 나타낸다. 그림에 나타난 기준이 되는 시 도표는 통계청에서 제공하는 월별 실업률을 바탕으로 산출된 것이다. 해당 기간 내에 24개월 이상 조사에 노출된 연동그룹만을 [그림 4-1]에 고려하였다. 연동그룹별 차이는 있으나 모두 고려된 24개의 연동그룹의 시 도표가 기준이 되는 시 도표와 큰 차이가 나타나지 않으며 무엇보다 연동그룹간의 체계적인 차이가 나타나지는 않는 것으로 파악된다. 다만 Panel1의 연동그룹 30과 Panel1의 연동그룹 31의 경우, 조사된 모든 시점에서 상대적으로 낮은 실업률을 보이고 있다.

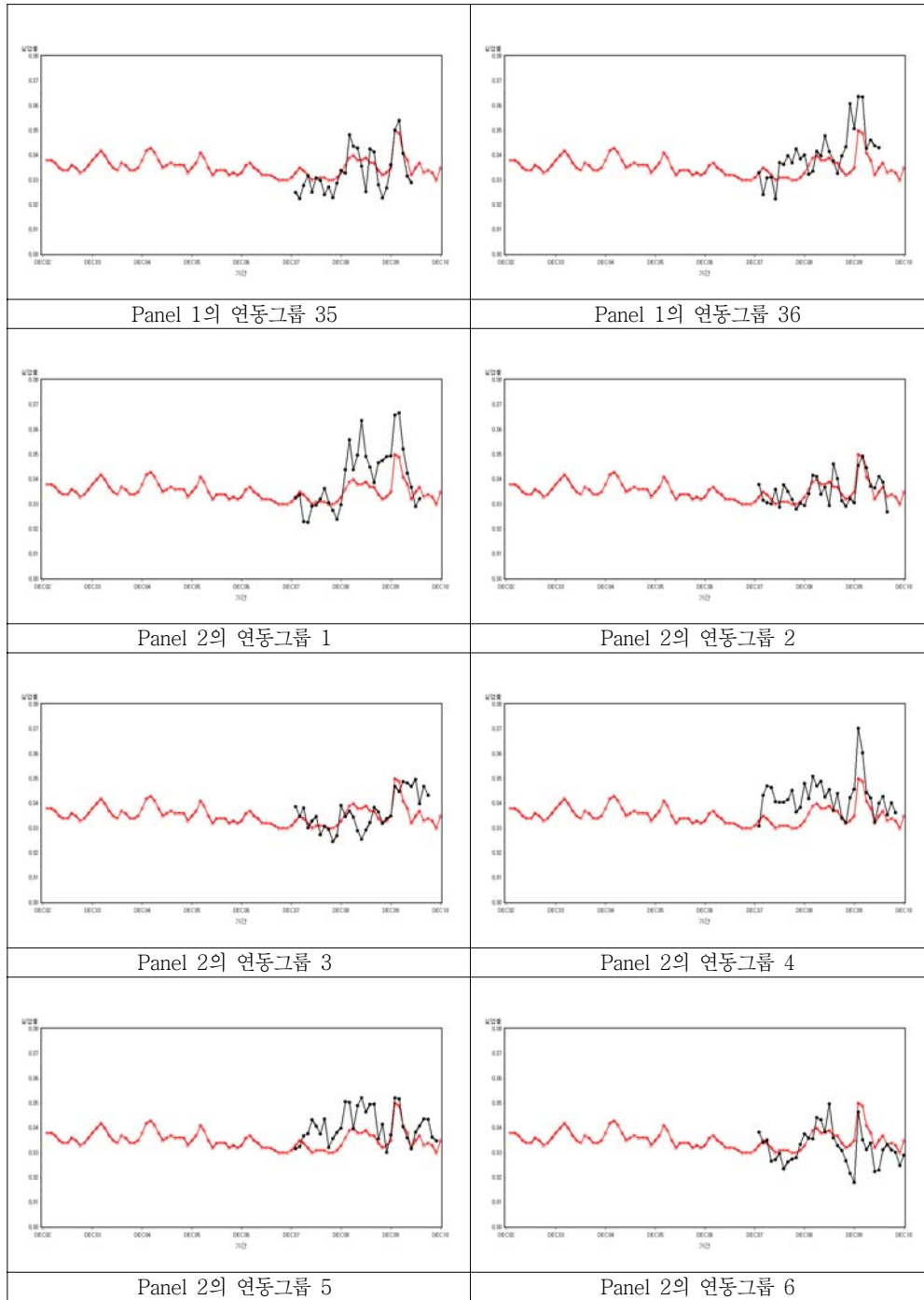
[그림 4-2]는 각 월별로 응답횟수별 실업률을 나타내고 있다. 조사 횟수 별(혹은 연동그룹별) 산출된 실업률이 모두 동일하지는 않으나 그 변화량이 대부분 크지 않고 무엇보다 응답횟수에 따른 체계적인 실업률의 변화가 발생하는 것으로는 보이지 않는다. 다만 몇 특정 월에서 36번째 조사에 노출된 그룹으로부터 계산된 실업률이 상대적으로 높게 나타나고 있다.

72개의 연동그룹별 2년간 실업률 자료를 검토해본 결과, 연동그룹별 체계적인 편향이 존재하는 것으로는 보기 힘들며 따라서 모집단 자료 혹은 동일 연동 그룹 내에서의 시계열 자료를 이용한 사후증화추정량 혹은 최량선형불편추정량의 사용이 적절하리라 여겨본다.

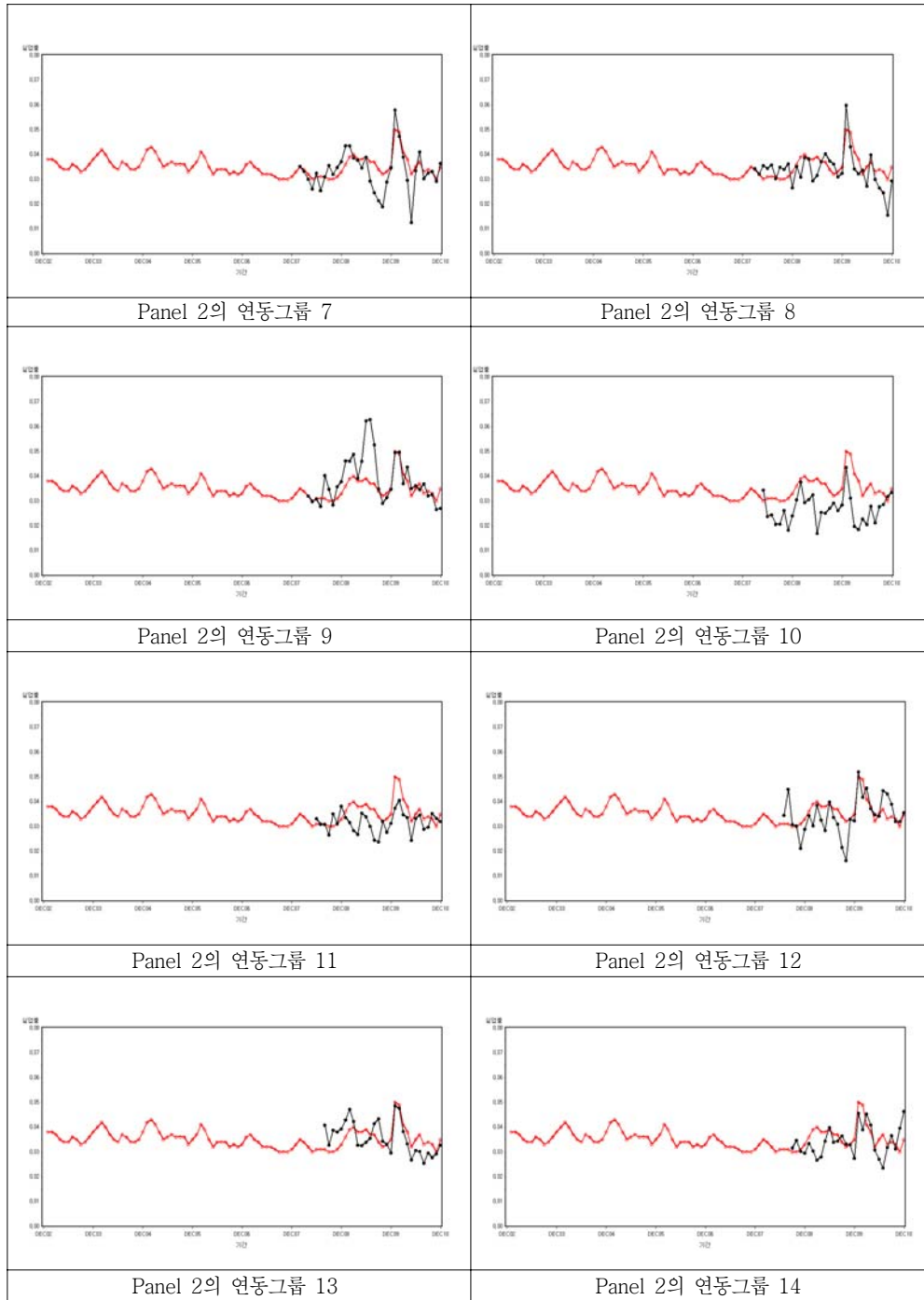
1) 매월 1/36씩 연동 교체되어 1개의 패널 당 36개씩 2개의 패널에서 총 72개의 연동그룹이 있음.



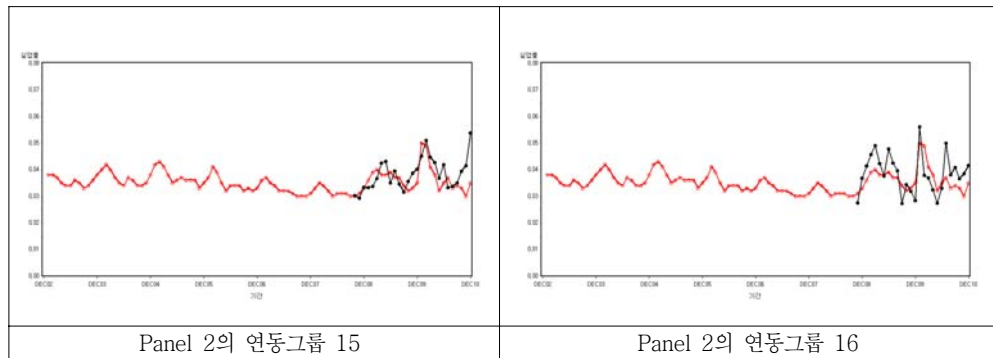
[그림 4-1] 연동그룹별 실업률 추정량분석



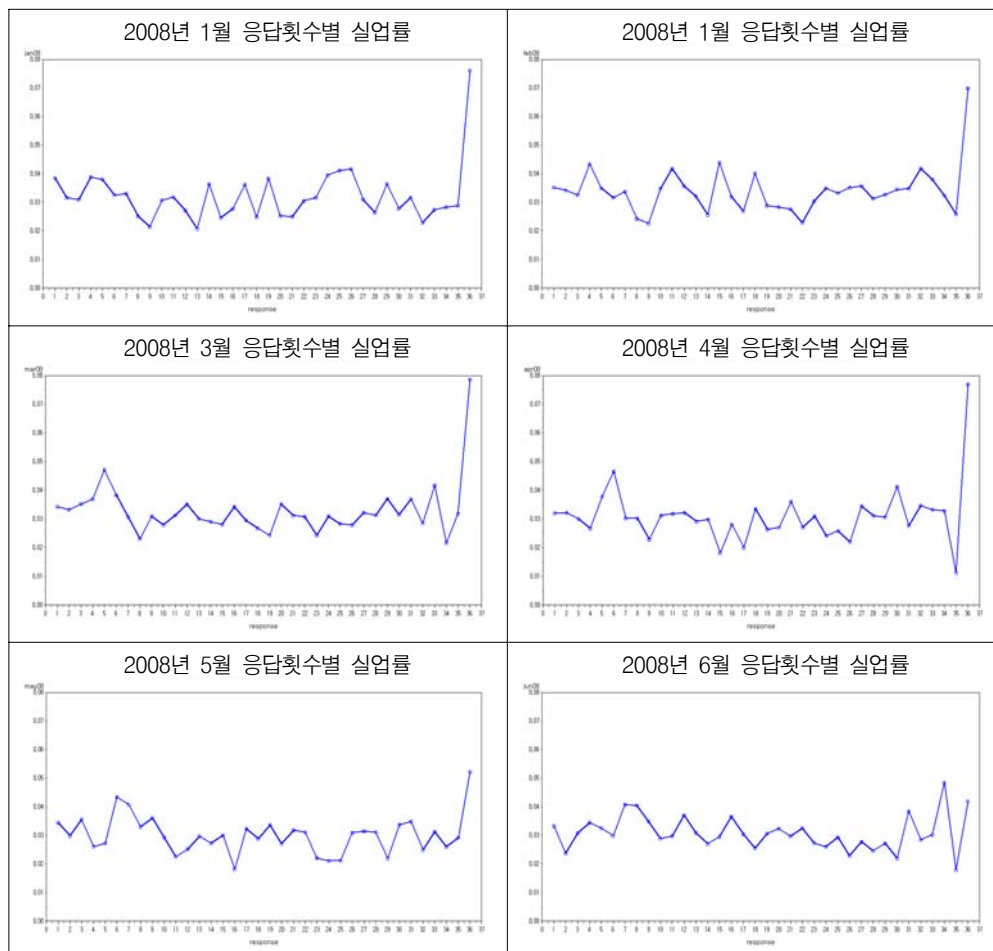
[그림 4-1] 연동그룹별 실업률 추정량분석 (계속)



[그림 4-1] 연동그룹별 실업률 추정량분석 (계속)



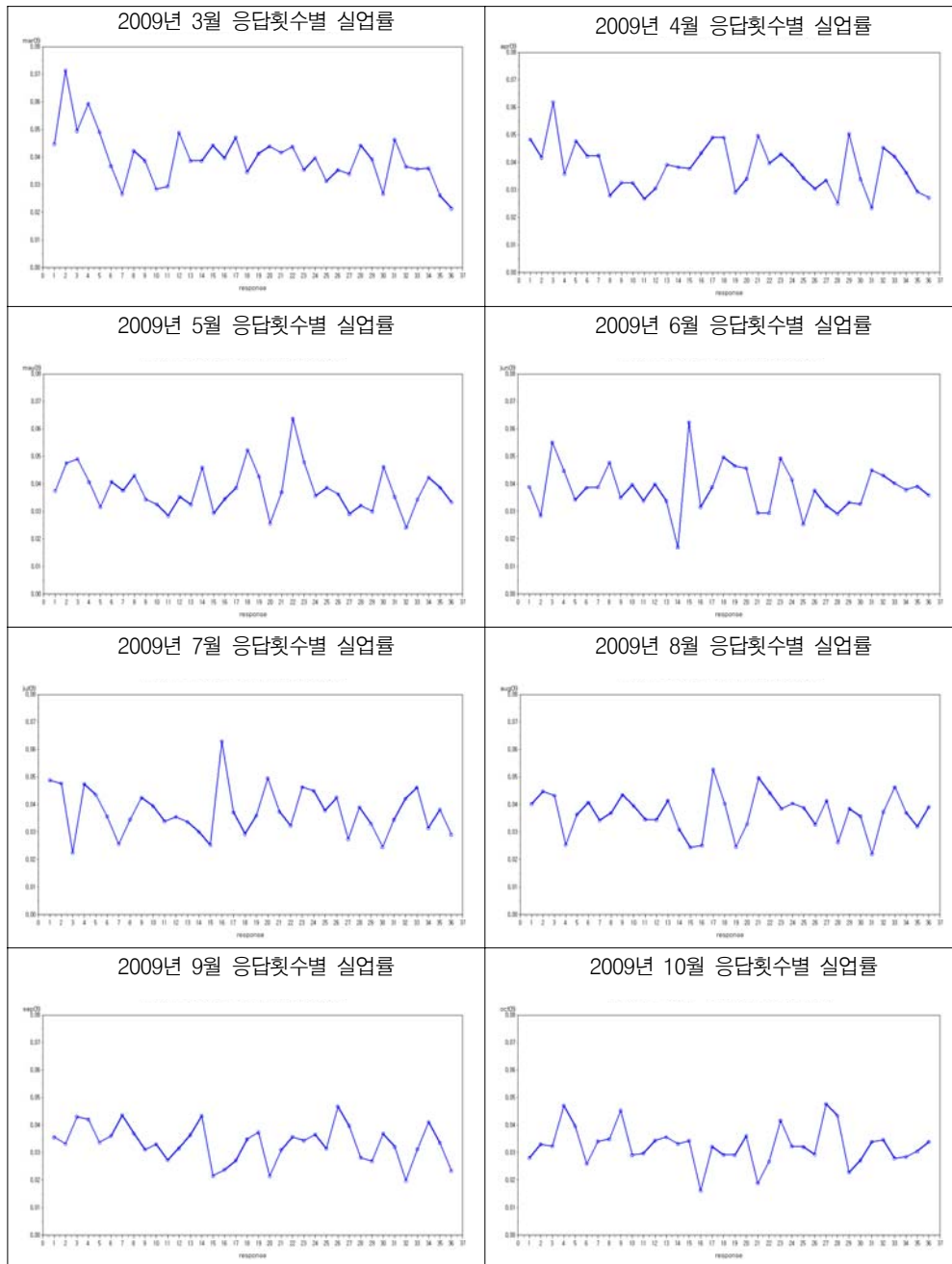
[그림 4-1] 연동그룹별 실업률 추정량분석 (계속)



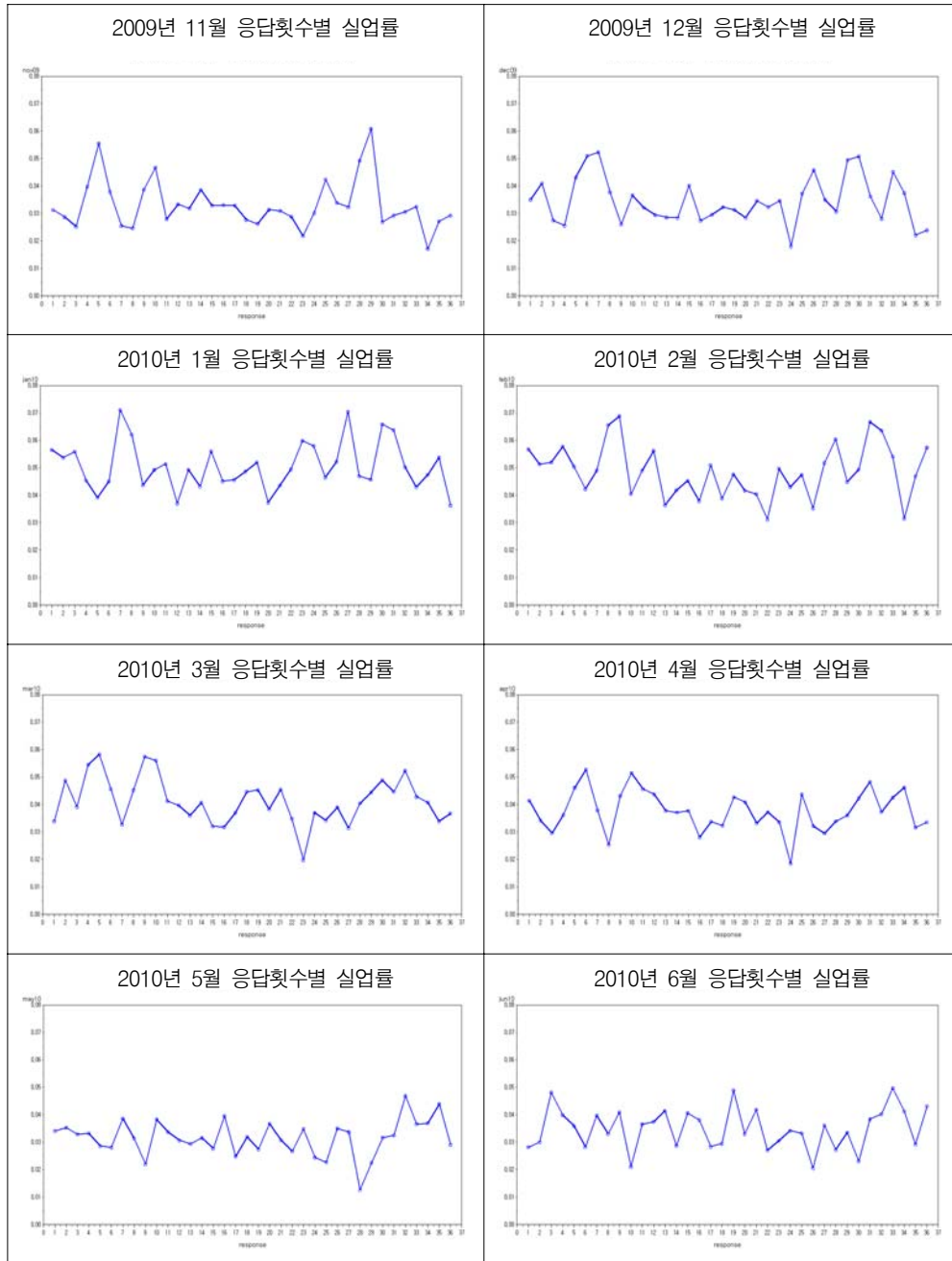
[그림 4-2] 응답횟수별 실업률 변동 추이



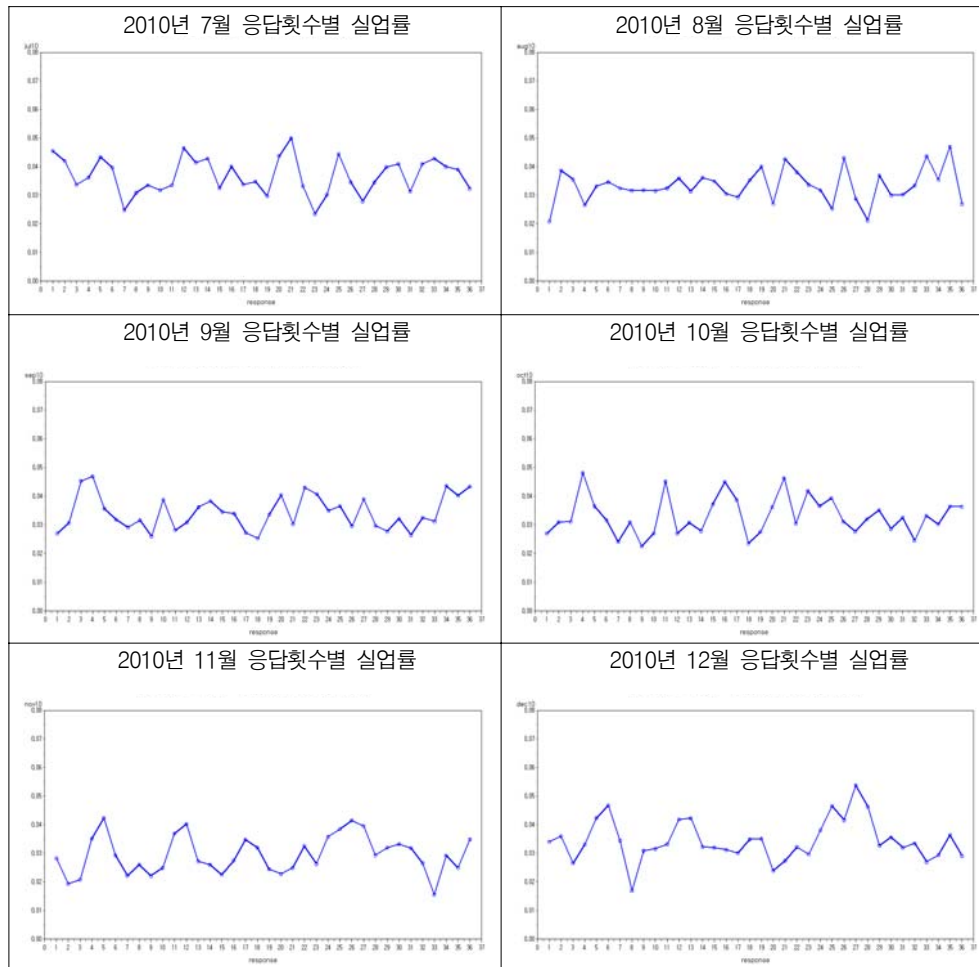
[그림 4-2] 응답횟수별 실업률 변동 추이 (계속)



[그림 4-2] 응답횟수별 실업률 변동 추이 (계속)



[그림 4-2] 응답횟수별 실업률 변동 추이 (계속)



[그림 4-2] 응답횟수별 실업률 변동 추이 (계속)

〈표 4-1〉 연동그룹별 실업률 추정량 분석 (계속)

[illegible]

〈표 4-1〉 연동그룹별 실업률 추정량 분석 (계속)

HSYYMM		21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
200801	unemp	0.0250	0.0253	0.0383	0.0248	0.0362	0.0277	0.0246	0.0364	0.0207	0.0271	0.0318	0.0308	0.0215	0.0251	0.0331
	N	1934	1879	1944	1917	1928	1891	1936	1888	2016	1954	2017	1980	1688	1724	1717
200802	unemp	0.0229	0.0276	0.0283	0.0288	0.0401	0.0269	0.0320	0.0439	0.0256	0.0321	0.0357	0.0418	0.0348	0.0226	0.0242
	N	1943	1873	1946	1924	1906	1885	1922	1888	2011	1955	2024	1980	1691	1744	1723
200803	unemp	0.0243	0.0307	0.0312	0.0351	0.0243	0.0267	0.0296	0.0342	0.0281	0.0290	0.0301	0.0351	0.0313	0.0279	0.0310
	N	1938	1829	1915	1895	1896	1864	1917	1874	1998	1918	1987	1967	1670	1716	1710
200804	unemp	0.0241	0.0309	0.0271	0.0360	0.0270	0.0263	0.0335	0.0200	0.0281	0.0181	0.0298	0.0291	0.0321	0.0318	0.0312
	N	1918	1825	1931	1897	1906	1865	1943	1886	2010	1918	2019	1973	1685	1705	1731
200805	unemp	0.0211	0.0211	0.0220	0.0311	0.0318	0.0270	0.0336	0.0288	0.0323	0.0181	0.0300	0.0273	0.0297	0.0252	0.0225
	N	1914	1823	1907	1899	1915	1856	1932	1877	1985	1911	2002	1977	1675	1701	1718
200806	unemp	0.0229	0.0293	0.0261	0.0273	0.0326	0.0297	0.0324	0.0306	0.0255	0.0304	0.0365	0.0295	0.0270	0.0309	0.0371
	N	1921	1851	1918	1918	1914	1875	1932	1894	1995	1932	2008	1982	1682	1700	1723
200807	unemp	0.0260	0.0323	0.0287	0.0241	0.0268	0.0254	0.0389	0.0278	0.0272	0.0252	0.0388	0.0203	0.0239	0.0300	0.0364
	N	1914	1866	1933	1913	1924	1873	1939	1905	2006	1935	2016	1983	1677	1687	1715
200808	unemp	0.0271	0.0400	0.0307	0.0204	0.0259	0.0261	0.0407	0.0215	0.0253	0.0181	0.0334	0.0173	0.0220	0.0242	0.0399
	N	1925	1855	1930	1915	1937	1865	1941	1915	2009	1934	1997	1964	1675	1681	1710
200809	unemp	0.0181	0.0271	0.0244	0.0253	0.0271	0.0237	0.0355	0.0334	0.0333	0.0143	0.0305	0.0256	0.0232	0.0273	0.0371
	N	1937	1861	1924	1897	1931	1855	1948	1907	1998	1922	1999	1966	1665	1675	1715
200810	unemp	0.0313	0.0327	0.0275	0.0338	0.0221	0.0287	0.0250	0.0336	0.0314	0.0316	0.0309	0.0266	0.0242	0.0230	0.0426
	N	1933	1860	1911	1875	1922	1841	1936	1906	1989	1914	1982	1946	1674	1683	1719
200811	unemp	0.0308	0.0270	0.0251	0.0366	0.0284	0.0262	0.0275	0.0270	0.0170	0.0268	0.0268	0.0265	0.0244	0.0288	0.0387
	N	1928	1853	1919	1876	1931	1846	1937	1899	1990	1921	1991	1935	1684	1697	1722
200812	unemp	0.0361	0.0300	0.0317	0.0364	0.0283	0.0289	0.0297	0.0431	0.0237	0.0253	0.0292	0.0329	0.0264	0.0340	0.0402
	N	1919	1862	1920	1873	1912	1856	1939	1922	1985	1928	1995	1946	1669	1701	1725
200901	unemp	0.0312	0.0347	0.0378	0.0348	0.0356	0.0338	0.0371	0.0482	0.0274	0.0326	0.0269	0.0320	0.0299	0.0329	0.0324
	N	1912	1859	1913	1860	1918	1840	1937	1923	1970	1921	2002	1939	1657	1708	1707
200902	unemp	0.0308	0.0447	0.0302	0.0417	0.0462	0.0304	0.0320	0.0485	0.0326	0.0322	0.0329	0.0287	0.0259	0.0484	0.0337
	N	1909	1852	1907	1863	1917	1855	1939	1916	1955	1913	1990	1943	1659	1693	1719
200903	unemp	0.0262	0.0360	0.0357	0.0366	0.0464	0.0268	0.0392	0.0442	0.0339	0.0353	0.0313	0.0397	0.0354	0.0437	0.0417
	N	1527	1827	1895	1844	1895	1871	1920	1887	1929	1894	1959	1927	1640	1665	1711
200904	unemp	0.0272	0.0293	0.0363	0.0422	0.0454	0.0234	0.0339	0.0504	0.0251	0.0335	0.0304	0.0343	0.0392	0.0430	0.0398
	N	1835	1817	1894	1828	1885	1849	1908	1892	1914	1884	1942	1902	1670	1657	1716
200905	unemp	.	0.0334	0.0387	0.0424	0.0344	0.0241	0.0353	0.0462	0.0300	0.0321	0.0291	0.0363	0.0387	0.0357	0.0479
	N	.	1822	1890	1809	1871	1854	1892	1874	1899	1884	1923	1891	1667	1640	1704
200906	unemp	.	0.0000	0.0359	0.0391	0.0379	0.0403	0.0431	0.0450	0.0327	0.0333	0.0291	0.0321	0.0376	0.0254	0.0416
	N	.	7	1891	1812	1865	1843	1890	1888	1910	1889	1937	1928	1671	1643	1703

〈표 4-1〉 연동그룹별 실업률 추정량 분석 (계속)

[illegible]

〈표 4-1〉 연동그룹별 실업률 추정량 분석 (계속)

HSYMM		Panel 1	Panel 2													
		36	1	2	3	4	5	6	7	8	9	10	11	12	13	14
200801	unemp	0.0326	0.0380	0.0388	0.0310	0.0317	0.0384	0.1245	0.0000	0.0000	0.1277	0.1353	0.0000	0.0000	0.0000	0.0000
	N	1671	2010	2073	1940	1936	1920	11	8	15	20	22	25	26	28	21
200802	unemp	0.0338	0.0317	0.0349	0.0434	0.0326	0.0342	0.0352	0.0695	0.0000	0.2293	0.1335	0.0000	0.0000	0.0000	0.0000
	N	1662	2001	2062	1967	1980	1978	1907	17	5	15	23	25	26	28	20
200803	unemp	0.0231	0.0307	0.0383	0.0471	0.0369	0.0352	0.0333	0.0343	0.0366	0.0784	0.0000	0.0000	0.0000	0.0000	0.0000
	N	1640	1958	2009	1934	1949	1984	1957	1864	23	8	5	20	27	23	20
200804	unemp	0.0228	0.0302	0.0302	0.0465	0.0378	0.0267	0.0301	0.0321	0.0320	0.1042	0.0000	0.0000	0.0000	0.0000	0.0000
	N	1639	1981	2007	1950	1981	1967	1964	1940	1884	11	4	8	21	17	20
200805	unemp	0.0293	0.0361	0.0330	0.0408	0.0434	0.0273	0.0261	0.0355	0.0298	0.0344	0.0000	0.0000	0.0000	0.0000	0.0000
	N	1631	1972	1986	1931	1965	1941	1942	1940	1900	1923	4	7	14	16	18
200806	unemp	0.0297	0.0289	0.0348	0.0405	0.0408	0.0298	0.0325	0.0344	0.0308	0.0238	0.0332	0.0000	0.0000	0.0000	0.0000
	N	1645	1984	1980	1949	1976	1947	1932	1957	1933	1975	1944	14	19	14	20
200807	unemp	0.0322	0.0379	0.0275	0.0406	0.0377	0.0236	0.0254	0.0358	0.0278	0.0244	0.0309	0.0345	0.0000	0.0000	0.0000
	N	1640	1981	1967	1951	1981	1969	1936	1972	1931	1985	1984	1941	18	19	18
200808	unemp	0.0364	0.0352	0.0308	0.0416	0.0437	0.0264	0.0310	0.0302	0.0403	0.0206	0.0308	0.0450	0.0408	0.0000	0.0000
	N	1628	1969	1960	1949	1967	1962	1937	1941	1951	1985	1990	1968	1869	15	8
200809	unemp	0.0307	0.0320	0.0296	0.0454	0.0322	0.0275	0.0356	0.0349	0.0348	0.0207	0.0266	0.0307	0.0327	0.0316	0.0000
	N	1621	1969	1956	1969	1971	1935	1920	1944	1945	1975	2011	1994	1895	1975	10
200810	unemp	0.0276	0.0281	0.0247	0.0366	0.0358	0.0281	0.0320	0.0340	0.0284	0.0262	0.0351	0.0303	0.0388	0.0347	0.0303
	N	1615	1983	1939	1959	1976	1931	1904	1944	1934	1969	2002	2002	1899	2030	1876
200811	unemp	0.0240	0.0305	0.0271	0.0385	0.0383	0.0335	0.0348	0.0362	0.0358	0.0183	0.0312	0.0212	0.0381	0.0302	0.0293
	N	1607	1985	1940	1964	1954	1942	1911	1946	1922	1970	1990	2017	1911	2034	1924
200812	unemp	0.0299	0.0296	0.0393	0.0481	0.0400	0.0377	0.0371	0.0266	0.0378	0.0240	0.0382	0.0289	0.0393	0.0295	0.0334
	N	1605	1966	1932	1972	1948	1936	1928	1961	1919	1972	1999	2026	1930	2034	1950
200901	unemp	0.0440	0.0343	0.0348	0.0420	0.0507	0.0359	0.0435	0.0352	0.0462	0.0305	0.0336	0.0344	0.0429	0.0334	0.0333
	N	1597	1955	1925	1954	1940	1944	1930	1964	1936	1967	2001	2021	1924	2019	1961
200902	unemp	0.0560	0.0417	0.0370	0.0510	0.0504	0.0357	0.0435	0.0309	0.0461	0.0376	0.0316	0.0303	0.0472	0.0304	0.0337
	N	1603	1966	1923	1954	1947	1936	1922	1975	1932	1971	2002	2011	1926	2008	1944
200903	unemp	0.0440	0.0413	0.0346	0.0471	0.0397	0.0443	0.0386	0.0387	0.0489	0.0294	0.0284	0.0386	0.0423	0.0267	0.0367
	N	1594	1943	1901	1927	1931	1921	1909	1957	1916	1921	1987	1980	1924	1981	1923
200904	unemp	0.0498	0.0341	0.0291	0.0490	0.0491	0.0434	0.0377	0.0383	0.0392	0.0306	0.0268	0.0326	0.0327	0.0280	0.0425
	N	1605	1948	1919	1924	1919	1923	1908	1951	1909	1930	1989	1969	1920	1975	1899
200905	unemp	0.0637	0.0370	0.0256	0.0427	0.0523	0.0386	0.0346	0.0294	0.0460	0.0325	0.0354	0.0284	0.0326	0.0344	0.0431
	N	1615	1945	1878	1927	1894	1913	1909	1926	1921	1933	1977	1980	1907	1992	1892
200906	unemp	0.0493	0.0295	0.0294	0.0457	0.0466	0.0498	0.0389	0.0316	0.0624	0.0169	0.0339	0.0399	0.0339	0.0398	0.0350
	N	1623	1942	1878	1917	1897	1919	1920	1934	1944	1946	1987	1990	1914	1984	1887

<표 4-1> 연동그룹별 실업률 추정량 분석 (계속)

HSYMM		Panel 1	Panel 2													
		36	1	2	3	4	5	6	7	8	9	10	11	12	13	14
200907	unemp	0.0450	0.0463	0.0324	0.0374	0.0496	0.0361	0.0293	0.0372	0.0629	0.0254	0.0301	0.0337	0.0355	0.0339	0.0395
	N	1635	1929	1907	1901	1902	1903	1928	1925	1935	1929	1980	1982	1910	1966	1896
200908	unemp	0.0388	0.0404	0.0385	0.0441	0.0497	0.0330	0.0246	0.0403	0.0527	0.0251	0.0245	0.0310	0.0414	0.0344	0.0346
	N	1637	1907	1908	1905	1910	1913	1935	1905	1927	1914	1962	1987	1892	1935	1895
200909	unemp	0.0468	0.0315	0.0366	0.0344	0.0357	0.0310	0.0214	0.0374	0.0349	0.0271	0.0238	0.0216	0.0434	0.0365	0.0316
	N	1612	1891	1908	1874	1915	1926	1919	1914	1913	1892	1958	1982	1891	1955	1868
200910	unemp	0.0477	0.0293	0.0321	0.0324	0.0416	0.0268	0.0189	0.0361	0.0291	0.0292	0.0322	0.0162	0.0343	0.0332	0.0356
	N	1594	1897	1897	1880	1903	1924	1906	1907	1893	1895	1968	1976	1874	1948	1853
200911	unemp	0.0492	0.0323	0.0340	0.0424	0.0302	0.0218	0.0289	0.0310	0.0314	0.0262	0.0277	0.0329	0.0331	0.0329	0.0387
	N	1598	1886	1896	1852	1893	1910	1914	1903	1885	1894	1963	1967	1872	1960	1869
200912	unemp	0.0495	0.0307	0.0350	0.0458	0.0373	0.0181	0.0346	0.0324	0.0347	0.0284	0.0313	0.0323	0.0296	0.0274	0.0402
	N	1604	1881	1925	1863	1894	1905	1915	1919	1887	1910	1984	1988	1870	1978	1891
201001	unemp	0.0659	0.0457	0.0470	0.0704	0.0522	0.0465	0.0580	0.0599	0.0495	0.0436	0.0374	0.0520	0.0486	0.0456	0.0451
	N	1601	1882	1927	1868	1899	1908	1910	1932	1885	1911	1990	1996	1886	1969	1883
201002	unemp	0.0668	0.0494	0.0449	0.0605	0.0518	0.0353	0.0474	0.0431	0.0498	0.0312	0.0405	0.0418	0.0476	0.0389	0.0510
	N	1609	1897	1922	1847	1908	1919	1899	1948	1883	1905	1983	1986	1868	1979	1882
201003	unemp	0.0523	0.0447	0.0489	0.0445	0.0405	0.0314	0.0390	0.0342	0.0370	0.0198	0.0348	0.0455	0.0383	0.0453	0.0447
	N	1584	1858	1881	1839	1890	1889	1885	1915	1860	1875	1942	1934	1846	1928	1844
201004	unemp	0.0426	0.0373	0.0483	0.0423	0.0361	0.0340	0.0296	0.0322	0.0437	0.0185	0.0336	0.0372	0.0333	0.0409	0.0427
	N	1589	1858	1869	1841	1888	1901	1885	1913	1864	1874	1946	1921	1828	1923	1838
201005	unemp	0.0369	0.0366	0.0469	0.0326	0.0317	0.0225	0.0126	0.0337	0.0350	0.0227	0.0244	0.0348	0.0268	0.0308	0.0368
	N	1593	1856	1868	1833	1915	1899	1857	1904	1858	1870	1919	1927	1832	1900	1826
201006	unemp	0.0292	0.0413	0.0498	0.0402	0.0384	0.0231	0.0335	0.0272	0.0362	0.0205	0.0333	0.0342	0.0306	0.0271	0.0419
	N	1580	1871	1861	1862	1911	1887	1862	1901	1857	1871	1935	1938	1840	1916	1840
201007	unemp	0.0323	0.0390	0.0400	0.0429	0.0409	0.0313	0.0410	0.0399	0.0346	0.0279	0.0346	0.0445	0.0302	0.0235	0.0333
	N	1595	1876	1871	1857	1908	1899	1880	1909	1859	1880	1940	1932	1839	1923	1825
201008	unemp	.	0.0270	0.0470	0.0355	0.0437	0.0334	0.0303	0.0300	0.0369	0.0212	0.0289	0.0431	0.0254	0.0318	0.0337
	N	.	1857	1870	1847	1895	1905	1876	1880	1855	1859	1950	1920	1845	1918	1822
201009	unemp	.	.	0.0434	0.0403	0.0436	0.0312	0.0325	0.0265	0.0321	0.0277	0.0297	0.0390	0.0296	0.0366	0.0350
	N	.	.	1836	1820	1867	1896	1859	1855	1854	1846	1929	1915	1827	1897	1809
201010	unemp	.	.	.	0.0364	0.0364	0.0303	0.0332	0.0246	0.0326	0.0286	0.0352	0.0320	0.0277	0.0312	0.0394
	N	.	.	.	1814	1868	1889	1870	1860	1863	1854	1938	1921	1837	1896	1833
201011	unemp	.	.	.	.	0.0349	0.0249	0.0292	0.0155	0.0266	0.0319	0.0333	0.0320	0.0293	0.0396	0.0415
	N	.	.	.	.	1858	1891	1872	1865	1857	1860	1959	1916	1851	1900	1823
201012	unemp	.	.	.	.	.	0.0291	0.0364	0.0294	0.0270	0.0335	0.0320	0.0356	0.0327	0.0464	0.0538
	N	.	.	.	.	.	1886	1867	1864	1871	1872	1952	1915	1849	1904	1822

〈표 4-1〉 연동그룹별 실업률 추정량 분석 (계속)

HSYMM		Panel 2														
		15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
200801	unemp	0.1179	0.1414	0.0000	0.0000	0.0000	0.0000	0.0000	0.1174	0.0000	0.1207	0.0000	0.0000	0.0000	0.0000	0.0000
	N	21	11	19	14	13	18	20	16	20	20	28	13	20	20	26
200802	unemp	0.1185	0.0000	0.0000	0.0000	0.0000	0.0000	0.1854	0.0000	0.0000	0.1474	0.0000	0.0000	0.0000	0.0000	0.0000
	N	21	13	21	14	13	17	20	15	20	18	28	13	19	20	26
200803	unemp	0.1138	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1712	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	21	12	21	15	13	17	19	15	21	15	27	15	18	20	26
200804	unemp	0.1026	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1582	0.0000	0.0000	0.0000	0.0000	0.0000
	N	21	16	21	15	13	16	19	16	21	15	27	15	18	20	26
200805	unemp	0.1438	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1662	0.0000	0.1205	0.0000	0.0000	0.0000
	N	17	14	21	15	13	17	19	16	21	15	26	15	18	20	26
200806	unemp	0.1338	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1829	0.0000	0.0000	0.1089	0.0000	0.0000
	N	25	16	18	15	14	17	19	15	22	15	24	15	18	20	27
200807	unemp	0.0718	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1770	0.0000	0.0000	0.0000	0.0898	0.0000
	N	25	17	18	15	15	17	19	17	22	15	24	15	18	20	27
200808	unemp	0.1158	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0879	0.0000
	N	18	19	21	15	15	16	19	17	22	15	24	15	18	20	27
200809	unemp	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1152	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	5	16	15	15	15	14	19	17	21	15	25	15	18	20	30
200810	unemp	0.0000	0.0000	0.0000	0.0000	0.0000	0.0706	0.1189	0.0979	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	10	11	18	22	19	28	20	16	21	14	23	15	18	20	30
200811	unemp	0.0275	0.0000	0.0000	0.1568	0.0000	0.0745	0.1305	0.0959	0.0000	0.1738	0.0622	0.0000	0.0000	0.0000	0.0000
	N	1860	9	13	22	16	28	20	16	21	14	24	15	19	20	30
200812	unemp	0.0368	0.0443	0.0000	0.1720	0.0000	0.0000	0.1345	0.1162	0.0000	0.0000	0.0000	0.0000	0.1812	0.0000	0.0665
	N	1923	1915	21	11	19	29	20	16	21	15	26	15	20	16	30
200901	unemp	0.0414	0.0510	0.0444	0.0000	0.0000	0.0000	0.1298	0.1293	0.0000	0.0000	0.0000	0.0000	0.1165	0.0000	0.0705
	N	1955	1956	1976	11	19	29	20	16	20	17	26	15	19	16	30
200902	unemp	0.0457	0.0532	0.0541	0.0467	0.0000	0.0000	0.1257	0.1312	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	N	1939	1987	2038	1866	19	25	19	16	20	18	25	15	19	17	30
200903	unemp	0.0491	0.0593	0.0495	0.0714	0.0447	0.1194	0.1188	0.0000	0.0000	0.1258	0.0000	0.0000	0.0000	0.0000	0.0000
	N	1929	1961	2030	1908	1891	13	13	16	20	18	25	15	19	16	27
200904	unemp	0.0423	0.0478	0.0358	0.0619	0.0418	0.0484	0.0000	0.0000	0.0000	0.1435	0.0000	0.0000	0.0997	0.0000	0.0000
	N	1929	1953	2018	1904	1946	1845	17	17	21	18	25	15	19	15	27
200905	unemp	0.0376	0.0408	0.0317	0.0407	0.0491	0.0476	0.0375	0.0000	0.0000	0.0740	0.0000	0.0000	0.1013	0.0000	0.0000
	N	1916	1943	1987	1904	1945	1863	1930	27	30	33	23	15	19	15	27
200906	unemp	0.0478	0.0389	0.0387	0.0343	0.0448	0.0551	0.0285	0.0389	0.1008	0.0000	0.0507	0.0000	0.1353	0.0000	0.0000
	N	1928	1958	1970	1915	1959	1891	1974	1936	32	33	25	15	19	12	28

〈표 4-1〉 연동그룹별 실업률 추정량 분석 (계속)

HSYMM		Panel 2														
		15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
200907	unemp	0.0425	0.0346	0.0257	0.0357	0.0436	0.0475	0.0225	0.0477	0.0489	0.0413	0.0000	0.1011	0.0000	0.0000	0.0000
	N	1939	1948	1982	1904	1962	1900	1971	1952	1966	32	25	13	19	14	29
200908	unemp	0.0396	0.0435	0.0370	0.0342	0.0407	0.0364	0.0254	0.0433	0.0448	0.0402	0.0853	0.1043	0.0000	0.1098	0.0000
	N	1930	1924	1968	1871	1943	1889	1937	1914	1978	1889	26	13	19	17	44
200909	unemp	0.0273	0.0331	0.0311	0.0370	0.0436	0.0361	0.0337	0.0421	0.0431	0.0332	0.0356	0.0000	0.0000	0.0000	0.0000
	N	1933	1928	1943	1859	1948	1874	1929	1887	1973	1903	1972	13	19	16	42
200910	unemp	0.0344	0.0297	0.0291	0.0454	0.0349	0.0341	0.0259	0.0397	0.0471	0.0324	0.0331	0.0281	0.0000	0.0000	0.0000
	N	1930	1935	1947	1867	1944	1845	1931	1876	1968	1913	1991	1892	20	16	42
200911	unemp	0.0319	0.0334	0.0280	0.0468	0.0387	0.0247	0.0255	0.0379	0.0556	0.0398	0.0253	0.0288	0.0313	0.0000	0.0000
	N	1936	1945	1967	1865	1942	1838	1935	1876	2019	1920	1999	1935	1905	16	42
200912	unemp	0.0284	0.0286	0.0295	0.0322	0.0367	0.0261	0.0378	0.0523	0.0510	0.0432	0.0256	0.0275	0.0410	0.0351	0.0671
	N	1928	1962	1979	1867	1943	1842	1950	1890	2027	1925	2010	1931	1941	1887	43
201001	unemp	0.0561	0.0432	0.0493	0.0370	0.0514	0.0493	0.0438	0.0621	0.0711	0.0451	0.0392	0.0454	0.0559	0.0537	0.0566
	N	1924	1956	1985	1869	1930	1835	1952	1901	2025	1938	2016	1955	1969	1911	1887
201002	unemp	0.0379	0.0453	0.0419	0.0363	0.0563	0.0490	0.0405	0.0688	0.0656	0.0491	0.0423	0.0505	0.0578	0.0520	0.0513
	N	1927	1953	1984	1867	1939	1820	1927	1893	2013	1948	1999	1953	1958	1918	1922
201003	unemp	0.0369	0.0318	0.0321	0.0407	0.0360	0.0397	0.0413	0.0561	0.0575	0.0453	0.0327	0.0457	0.0583	0.0545	0.0391
	N	1907	1924	1960	1836	1916	1805	1922	1857	1970	1915	1968	1937	1926	1909	1897
201004	unemp	0.0324	0.0338	0.0281	0.0377	0.0371	0.0378	0.0437	0.0456	0.0515	0.0432	0.0253	0.0379	0.0527	0.0462	0.0362
	N	1892	1922	1936	1845	1896	1807	1923	1859	1958	1887	1961	1924	1924	1916	1890
201005	unemp	0.0274	0.0320	0.0248	0.0395	0.0277	0.0316	0.0294	0.0307	0.0337	0.0383	0.0220	0.0314	0.0386	0.0281	0.0286
	N	1884	1909	1917	1846	1905	1814	1907	1846	1962	1867	1945	1909	1896	1905	1878
201006	unemp	0.0330	0.0489	0.0295	0.0284	0.0382	0.0406	0.0287	0.0414	0.0375	0.0365	0.0211	0.0409	0.0330	0.0398	0.0283
	N	1910	1926	1929	1851	1922	1834	1912	1864	1974	1882	1926	1924	1888	1891	1885
201007	unemp	0.0500	0.0438	0.0297	0.0348	0.0338	0.0400	0.0326	0.0429	0.0415	0.0465	0.0336	0.0318	0.0335	0.0309	0.0248
	N	1917	1948	1923	1840	1935	1844	1912	1868	1968	1914	1923	1929	1886	1916	1880
201008	unemp	0.0381	0.0427	0.0270	0.0400	0.0354	0.0293	0.0305	0.0349	0.0362	0.0314	0.0360	0.0325	0.0316	0.0319	0.0317
	N	1913	1940	1937	1834	1920	1840	1900	1869	1957	1893	1908	1909	1883	1914	1859
201009	unemp	0.0408	0.0430	0.0303	0.0405	0.0337	0.0254	0.0273	0.0339	0.0346	0.0383	0.0363	0.0310	0.0282	0.0388	0.0260
	N	1889	1923	1918	1840	1910	1822	1893	1846	1956	1889	1875	1899	1869	1896	1854
201010	unemp	0.0366	0.0418	0.0304	0.0463	0.0362	0.0276	0.0235	0.0387	0.0450	0.0372	0.0279	0.0308	0.0271	0.0451	0.0271
	N	1886	1926	1905	1822	1888	1827	1884	1864	1968	1906	1886	1906	1888	1907	1866
201011	unemp	0.0385	0.0358	0.0263	0.0326	0.0249	0.0228	0.0245	0.0320	0.0348	0.0275	0.0225	0.0260	0.0272	0.0402	0.0369
	N	1885	1930	1894	1812	1890	1848	1877	1883	1950	1908	1880	1912	1881	1910	1866
201012	unemp	0.0416	0.0466	0.0382	0.0297	0.0322	0.0274	0.0240	0.0351	0.0349	0.0301	0.0312	0.0320	0.0323	0.0423	0.0419
	N	1887	1939	1914	1818	1896	1841	1893	1870	1945	1894	1870	1915	1853	1918	1874

〈표 4-1〉 연동그룹별 실업률 추정량 분석 (계속)

HSYYMM		Panel 2							Panel 3			
		30	31	32	33	34	35	36	1	2	3	4
200801	unemp	0.0000	0.0000	0.1198	0.0110	0.0525	0.0217	0.0138	.	.	.	.
	N	10	12	12	293	271	288	261	.	.	.	.
200802	unemp	0.0000	0.0000	0.0000	0.0161	0.0453	0.0256	0.0072	.	.	.	.
	N	18	21	21	291	270	293	260	.	.	.	.
200803	unemp	0.0000	0.0000	0.0000	0.0331	0.0261	0.0313	0.0223	.	.	.	.
	N	19	21	20	284	268	290	249	.	.	.	.
200804	unemp	0.0000	0.0000	0.0000	0.0268	0.0210	0.0317	0.0206	.	.	.	.
	N	19	21	21	277	271	290	259	.	.	.	.
200805	unemp	0.0597	0.0000	0.0000	0.0324	0.0255	0.0060	0.0400	.	.	.	.
	N	19	21	22	279	273	288	255	.	.	.	.
200806	unemp	0.0000	0.0000	0.0000	0.0318	0.0256	0.0147	0.0305	.	.	.	.
	N	20	21	23	282	272	291	252	.	.	.	.
200807	unemp	0.0000	0.0000	0.0000	0.0105	0.0366	0.0199	0.0066	.	.	.	.
	N	20	21	23	283	268	294	252	.	.	.	.
200808	unemp	0.0000	0.0000	0.0000	0.0051	0.0262	0.0144	0.0128	.	.	.	.
	N	18	21	23	276	271	290	258	.	.	.	.
200809	unemp	0.0000	0.0879	0.0000	0.0178	0.0079	0.0348	0.0144	.	.	.	.
	N	18	21	22	281	269	286	259	.	.	.	.
200810	unemp	0.0000	0.0000	0.0000	0.0208	0.0202	0.0172	0.0167	.	.	.	.
	N	18	21	22	273	265	285	256	.	.	.	.
200811	unemp	0.0000	0.0000	0.0000	0.0256	0.0203	0.0213	0.0311	.	.	.	.
	N	18	21	22	273	264	280	258	.	.	.	.
200812	unemp	0.0000	0.0000	0.0000	0.0321	0.0154	0.0221	0.0496	.	.	.	.
	N	20	20	22	270	266	278	251	.	.	.	.
200901	unemp	0.0000	0.0000	0.0000	0.0320	0.0161	0.0076	0.0462	.	.	.	.
	N	20	20	22	273	266	270	251	.	.	.	.
200902	unemp	0.0000	0.0000	0.1248	0.0315	0.0365	0.0073	0.0569	.	.	.	.
	N	20	19	22	272	266	270	252	.	.	.	.
200903	unemp	0.0000	0.0000	0.1202	0.0427	0.0360	0.0414	0.0653	.	.	.	.
	N	20	21	22	265	253	267	247	.	.	.	.
200904	unemp	0.0000	0.0524	0.0000	0.0385	0.0271	0.0274	0.0779	.	.	.	.
	N	21	21	22	266	261	267	249	.	.	.	.
200905	unemp	0.0000	0.0519	0.0000	0.0332	0.0349	0.0512	0.1023	.	.	.	.
	N	21	21	22	264	262	265	247	.	.	.	.
200906	unemp	0.0000	0.0517	0.0000	0.0193	0.0533	0.0441	0.0708	.	.	.	.
	N	21	21	22	273	268	260	249	.	.	.	.

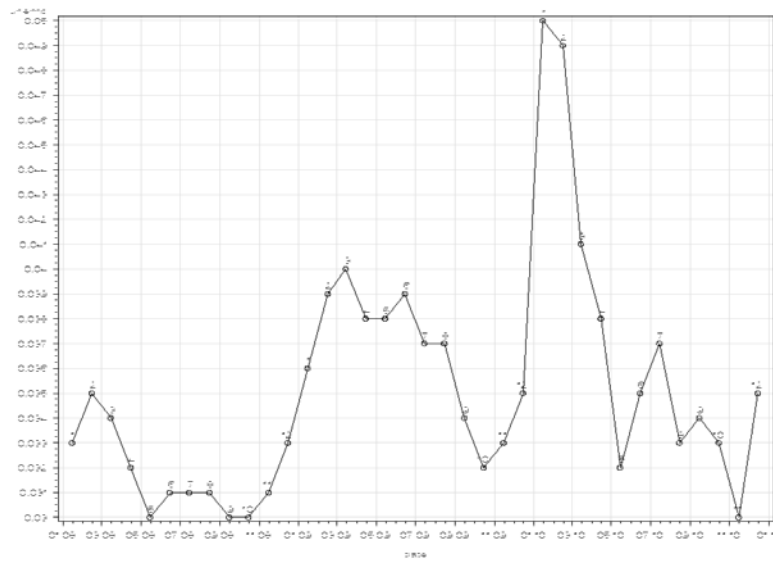
〈표 4-1〉 연동그룹별 실업률 추정량 분석 (계속)

HSYYMM		Panel 2							Panel 3			
		30	31	32	33	34	35	36	1	2	3	4
200907	unemp	0.0000	0.0745	0.0000	0.0345	0.0456	0.0175	0.0616	.	.	.	.
	N	21	21	22	276	265	266	252	.	.	.	.
200908	unemp	0.0000	0.0176	0.1739	0.0293	0.0283	0.0359	0.0567	.	.	.	.
	N	26	35	31	279	267	257	243	.	.	.	.
200909	unemp	0.1363	0.1078	0.2750	0.0242	0.0295	0.0238	0.0502	.	.	.	.
	N	27	34	31	275	267	251	242	.	.	.	.
200910	unemp	0.1769	0.0813	0.0000	0.0056	0.0156	0.0289	0.0405	.	.	.	.
	N	22	34	31	269	271	258	239	.	.	.	.
200911	unemp	0.1317	0.0807	0.0000	0.0425	0.0348	0.0483	0.0443	.	.	.	.
	N	28	34	31	268	269	251	246	.	.	.	.
200912	unemp	0.1368	0.0942	0.0642	0.0403	0.0123	0.0514	0.0497	.	.	.	.
	N	28	36	31	264	268	258	244	.	.	.	.
201001	unemp	0.1439	0.0977	0.0675	0.0225	0.0646	0.0457	0.0488	.	.	.	.
	N	28	36	31	263	267	259	242	.	.	.	.
201002	unemp	0.0567	0.0000	0.0654	0.0444	0.0594	0.0434	0.0396	.	.	.	.
	N	1910	36	32	255	269	258	244	.	.	.	.
201003	unemp	0.0488	0.0339	0.0000	0.0190	0.0495	0.0570	0.0217	.	.	.	.
	N	1900	1817	29	250	269	260	242	.	.	.	.
201004	unemp	0.0296	0.0342	0.0415	0.0059	0.0307	0.0386	0.0301	.	.	.	.
	N	1913	1851	1820	246	263	247	243	.	.	.	.
201005	unemp	0.0333	0.0329	0.0354	0.0341	0.0268	0.0463	0.0290	.	.	.	.
	N	1904	1821	1826	1599	264	248	239	.	.	.	.
201006	unemp	0.0360	0.0399	0.0482	0.0300	0.0282	0.0198	0.0188	.	.	.	.
	N	1905	1815	1830	1633	1657	248	250	.	.	.	.
201007	unemp	0.0398	0.0434	0.0362	0.0338	0.0420	0.0455	0.0340	.	.	.	.
	N	1891	1820	1818	1613	1718	1680	253	.	.	.	.
201008	unemp	0.0325	0.0346	0.0332	0.0266	0.0356	0.0387	0.0208	.	.	.	.
	N	1867	1802	1812	1616	1681	1704	1635	.	.	.	.
201009	unemp	0.0316	0.0292	0.0319	0.0357	0.0469	0.0454	0.0307	0.0271	.	.	.
	N	1857	1808	1796	1593	1642	1693	1649	1850	.	.	.
201010	unemp	0.0225	0.0309	0.0241	0.0317	0.0365	0.0481	0.0312	0.0310	0.0271	.	.
	N	1859	1817	1799	1610	1649	1710	1644	1890	1900	.	.
201011	unemp	0.0249	0.0221	0.0261	0.0222	0.0293	0.0424	0.0352	0.0208	0.0194	0.0283	.
	N	1867	1829	1822	1609	1638	1707	1657	1909	1930	1876	.
201012	unemp	0.0332	0.0316	0.0309	0.0169	0.0345	0.0468	0.0424	0.0331	0.0266	0.0360	0.0341
	N	1865	1818	1816	1611	1639	1717	1668	1915	1937	1916	1788

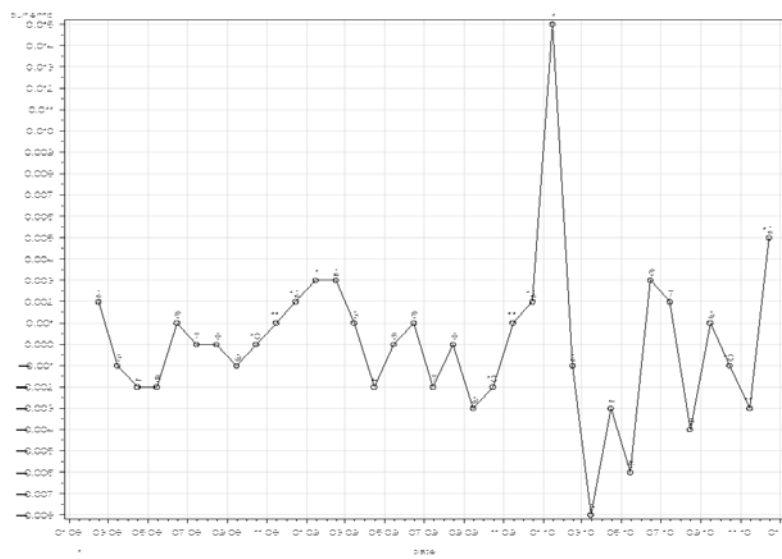
3. 최량선형불편추정량 사용을 위한 시계열 모형 적합

본 절에서는 현행 사용하고 있는 회귀추정량의 하나인 사후충화추정량 대신 사용할 수 있는 최량선형불편추정량의 정의에 있어 필수적인 각 연동그룹별 조사횟수에 따른 추정량들의 공분산 혹은 상관 구조의 모형화 방안을 살펴보도록 한다. 이는 최량선형불편 추정량의 계산을 위해서는 산출된 각 연동그룹별, 월별 추정량의 공분산 행렬의 추정이 필요하며 매월 계산될 수 있는 추정량들의 표본 분산-공분산 행렬을 사용하는 것보다 전체 자료에 대한 모형적합을 통해 통계적으로 타당한 모형을 정의하고 이를 바탕으로 최량선형불편추정량을 산출하는 것이 보다 안정적인 추정량을 제공할 수 있으리라 예측되기 때문이다. 모형적합을 위하여 기본적으로 서로 다른 연동그룹으로부터 산출된 추정량들은 그 조사 대상자가 서로 다름으로 독립이나 동일 연동그룹의 각 월별 추정량들은 서로 연관되어 있다고 가정한다. 이러한 가정 하에서 먼저는 전체 자료에 대한 시계열 적합 결과를 통해 시간의 경과에 따른 추정량들의 적절한 시계열 모형을 설정하고 이 결과를 각 연동그룹별 추정량자료와 비교하여 고려된 시계열 모형이 각 연동그룹별 모형적합을 위해 적합한지를 파악하였다.

전 장에서 고려한 추정량들의 시계열 모형 검토를 위해서 2008년부터 2010년 12월까지의 통계청 제공 실업률 자료를 이용하여 시계열 분석을 실시하였다. 시계열 자료 분석을 위해서 먼저 시간의 흐름에 따른 실업률을 보기 위하여 [그림 4-3]의 원자료 시 도표를 작성하였다. 2010년 1월과 2월의 실업률이 비정상적으로 커지는 경우를 제외하고는 전체적으로 정상성 조건 중 하나인 등분산성의 가정은 만족하는 것으로 판단된다. 유의한 선형 추세는 보이지 않으나 3년간 각 년도 별 실업률이 증가 추세에 있다고 판단되어 차분된 자료를 이용하여 시계열 분석을 실시하였다. [그림 4-4]는 1차 차분이 이루어진 후의 시 도표를 나타내고 있다. 일반적으로 실업률과 같은 노동 시장 관련 통계들은 1년 단위의 계절효과(seasonality)를 갖고 있는 것으로 알려져 있다. 그러나 두 시 도표를 볼 때 이러한 계절효과는 존재하지 않는 것으로 판단된다. 특별히 [그림 4-5]와 [그림 4-6]의 시계열 자료의 ACF와 PACF를 볼 때 본 3년의 자료에 일반적인 1년 단위의 계절 효과는 없는 것으로 판단된다. 이는 시계열 분석을 위한 자료가 계절효과를 파악할 수 있을 정도로 크지 않고 3년간의 노동시장이 상황이 불안정하였기 때문일 수도 있다.



[그림 4-3] 실업률 자료(y_t)의 시 도표



[그림 4-4] 실업률 자료(y_t)의 시 도표

Autocorrelations																								
Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	Std Error
0	0.00001331	1.00000																						0
1	1.49378E-6	0.11222																						0.169031
2	-2.6557E-6	-0.19951																						0.171146
3	-2.375E-6	-0.17842																						0.177667
4	-5.5481E-6	-0.41681																						0.182715
5	2.30402E-6	0.17309																						0.208115
6	1.73577E-6	0.13040																						0.212189
7	-1.3737E-6	-0.10320																						0.214466
8	5.00478E-7	0.03760																						0.215880
9	-3.3902E-6	-0.25469																						0.216067
10	-8.3032E-7	-0.06238																						0.224481
11	3.8855E-6	0.29190																						0.224976
12	1.80051E-6	0.13526																						0.235549
13	1.15144E-6	0.08650																						0.237758
14	-1.1882E-6	-0.09327																						0.238655
15	-1.6806E-6	-0.12526																						0.239607
16	-4.2843E-7	-0.03219																						0.241501
17	-3.1261E-7	-0.02348																						0.241623
18	4.8077E-7	0.03612																						0.241688
19	3.1088E-7	0.02336																						0.241843
20	-1.5227E-6	-0.11439																						0.241907
21	-1.3864E-6	-0.10416																						0.243448
22	1.88152E-7	0.01414																						0.244718
23	1.68194E-6	0.12636																						0.244741
24	9.68373E-7	0.07275																						0.246598

[그림 4-5] 실업률 자료(∇y_t)의 ACF

Partial Autocorrelations																							
Lag	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
1	0.11222																						
2	-0.21481																						
3	-0.13497																						
4	-0.45604																						
5	0.23918																						
6	-0.18854																						
7	-0.14171																						
8	-0.12403																						
9	-0.22527																						
10	-0.14328																						
11	0.08294																						
12	0.03990																						
13	-0.08305																						
14	-0.02133																						
15	0.15236																						
16	-0.13554																						
17	-0.01936																						
18	-0.00681																						
19	0.02751																						
20	-0.19369																						
21	-0.05977																						
22	-0.05130																						
23	-0.04325																						
24	-0.17014																						

[그림 4-6] 실업률 자료(∇y_t)의 PACF〈표 4-2〉 실업률 자료(∇y_t)의 백색잡음 검정

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	12.64	6	0.0491	0.112	-0.2	-0.178	-0.417	0.173	0.13
12	22.26	12	0.0347	-0.103	0.038	-0.255	-0.062	0.292	0.135
18	24.43	18	0.1413	0.087	-0.089	-0.126	-0.032	-0.023	0.036
24	28.98	24	0.221	0.023	-0.114	-0.104	0.014	0.126	0.073

[그림 4-5], [그림 4-6] 그리고 <표 4-2>의 포트만토 검정결과를 이용하여 ARIMA(4,1,0), ARIMA(0,1,4) 그리고 ARIMA(4,1,4)의 3가지 모형을 고려하였다. <표 4-3> 각 모형을 통하여 적합된 결과를 나타내고 있다. ARIMA(0,1,4) 모형이 최소의 AIC와 SBC 값을 가지고 있으므로 최소 AIC기준에 따라 만족스럽다고 할 수 있으나 그 차이가 거의 미미하고 또한 ACF를 통해 볼 때 ARIMA(4, 1, 0)모형이 더 적절한 모형으로 간주된다. 일반적으로 안정적인 ARMA 모형을 적합하기 위해서는 자료수가 어느 정도 이상인 것이 바람직하며 자료수가 약 50개 미만인 경우가 급적 AR모형의 적합을 시도하는 것이 좋은 것으로 알려져 있다.(박유성·김기환 2002).

<표 4-3> 세 가지 모형에서의 AIC와 SBC값 (ML추정방법)

모형	AIC	SBC
ARIMA(4, 1, 0)	-296.318	-293.208
ARIMA(0, 1, 4)	-296.327	-293.217
ARIMA(4, 1, 4)	-294.852	-290.186

<표 4-4>와 <표 4-5>는 ARIMA(4, 1, 0)의 적합결과를 나타내고 있다. 적합 후 잔차들의 자기상관 여부를 포트만토 검정방법을 통하여 평가하였다. 완전한 적합이라고 평가하기는 어려우나 $k=6$ 인 검정 결과 원자료의 검정결과와는 달리 모든 경우의 유의확률이 0.05보다 큼을 확인할 수 있다. 모형적합 결과 t 시점의 실업률을 예측하기 위한 모형은 다음과 같이 추정된다.

$$(1 + 0.41370B^4)\nabla y_t = a_t$$

$$\Rightarrow y_t = y_{t-1} - 0.41(y_{t-4} - y_{t-5}) + a_t$$

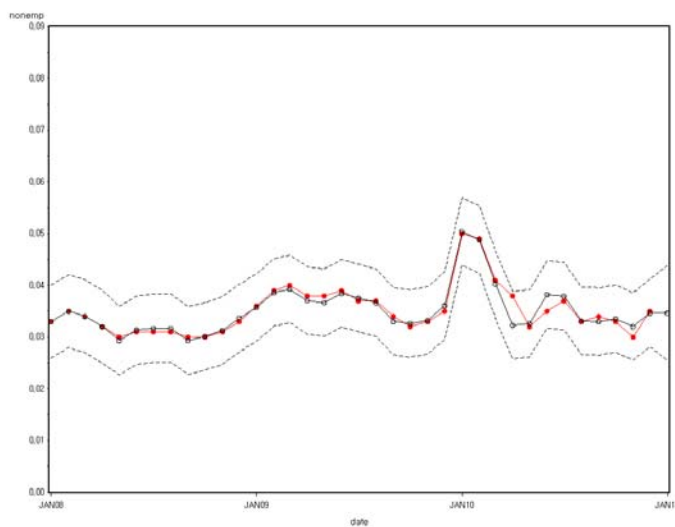
<표 4-4> 상수항이 없는 ∇y_t 의 모형에서 모수 추정치와 적합도 통계치

Parameter	Estimate	Standard Error	t Value	Approx	Lag
				Pr > t	
AR1,1	-0.4137	0.15858	-2.61	0.0091	4

〈표 4-5〉 ARI(4,1)에 대한 포트만토 검정치

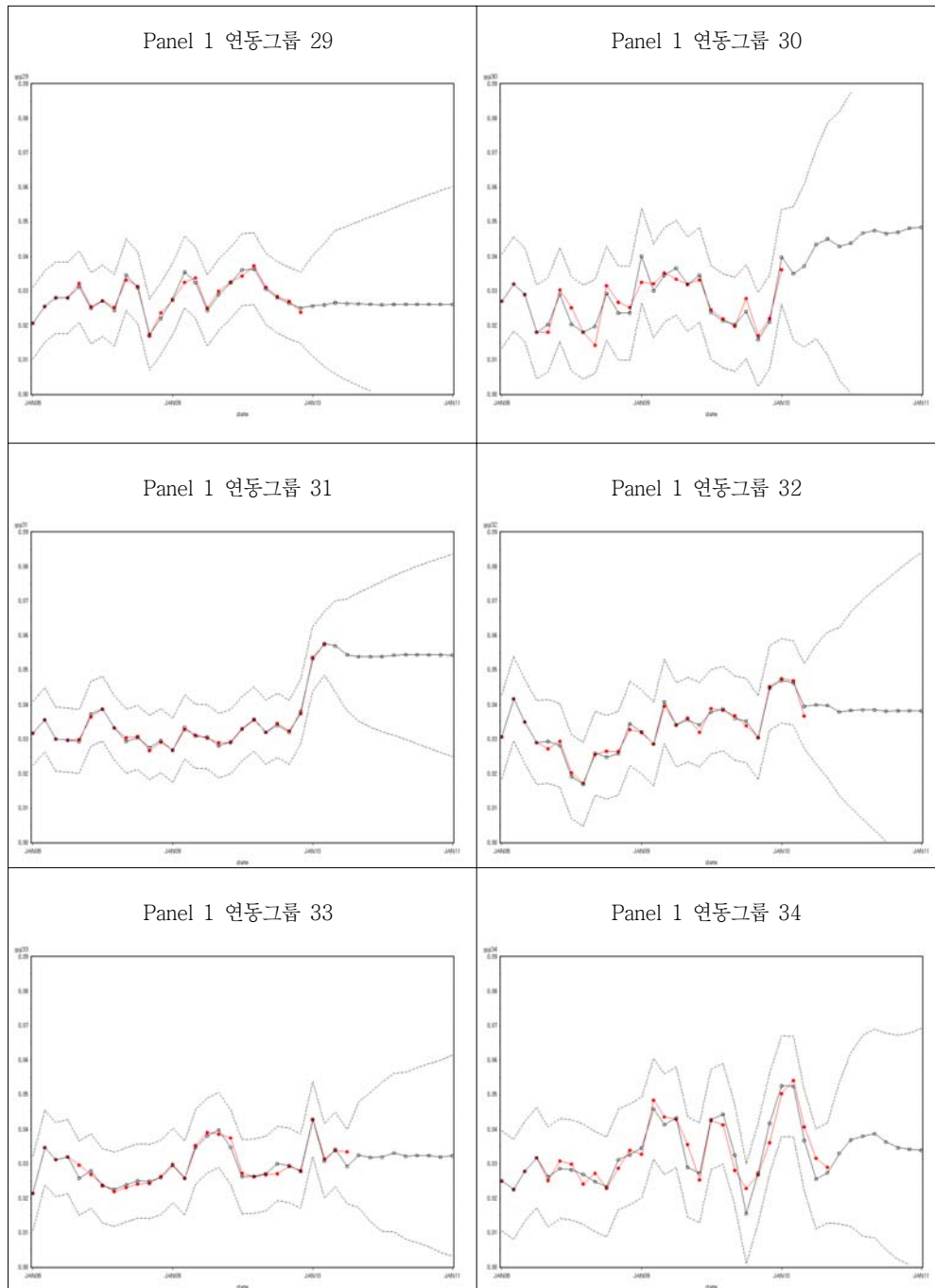
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	9.41	5	0.0937	0.161	-0.313	-0.246	-0.097	0.195	0.047
12	18.8	11	0.0647	-0.101	-0.059	-0.213	-0.072	0.293	0.168
18	20.18	17	0.2649	0.032	-0.084	-0.089	-0.048	-0.053	0.013
24	26.75	23	0.267	0.046	-0.136	-0.135	0.015	0.155	0.064

[그림 4-7]은 36개월의 실업률 자료를 통해 추정된 모형 ARIMA(4,1,0)을 통해 예측된 값과 대응 신뢰구간을 나타내고 있다. 그림을 통해 볼 때 적용 모형의 적합도와 예측력은 매우 높은 것으로 판단된다.

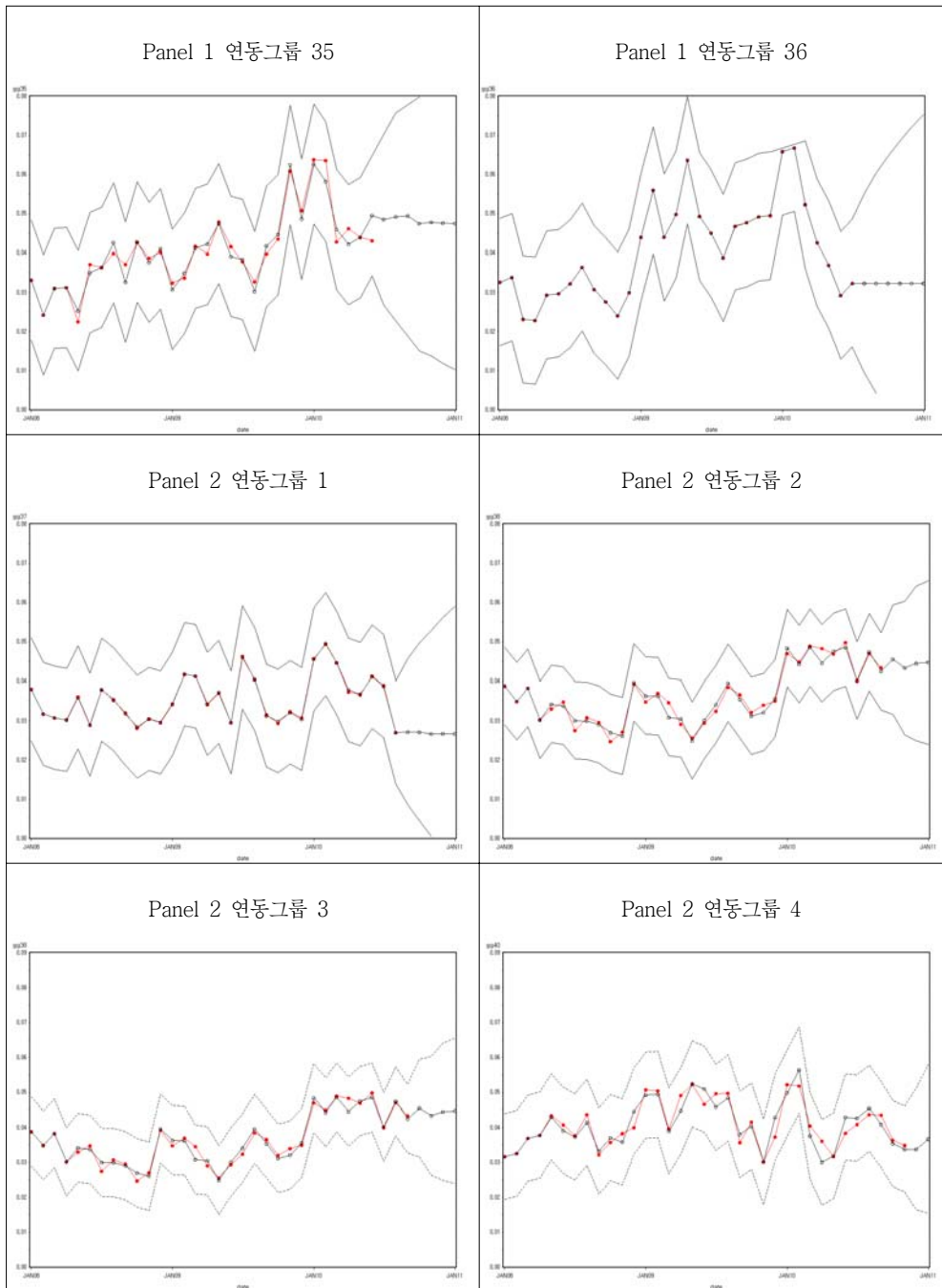


[그림 4-7] 실업률 자료의 시 도표 - 예측값 및 신뢰구간

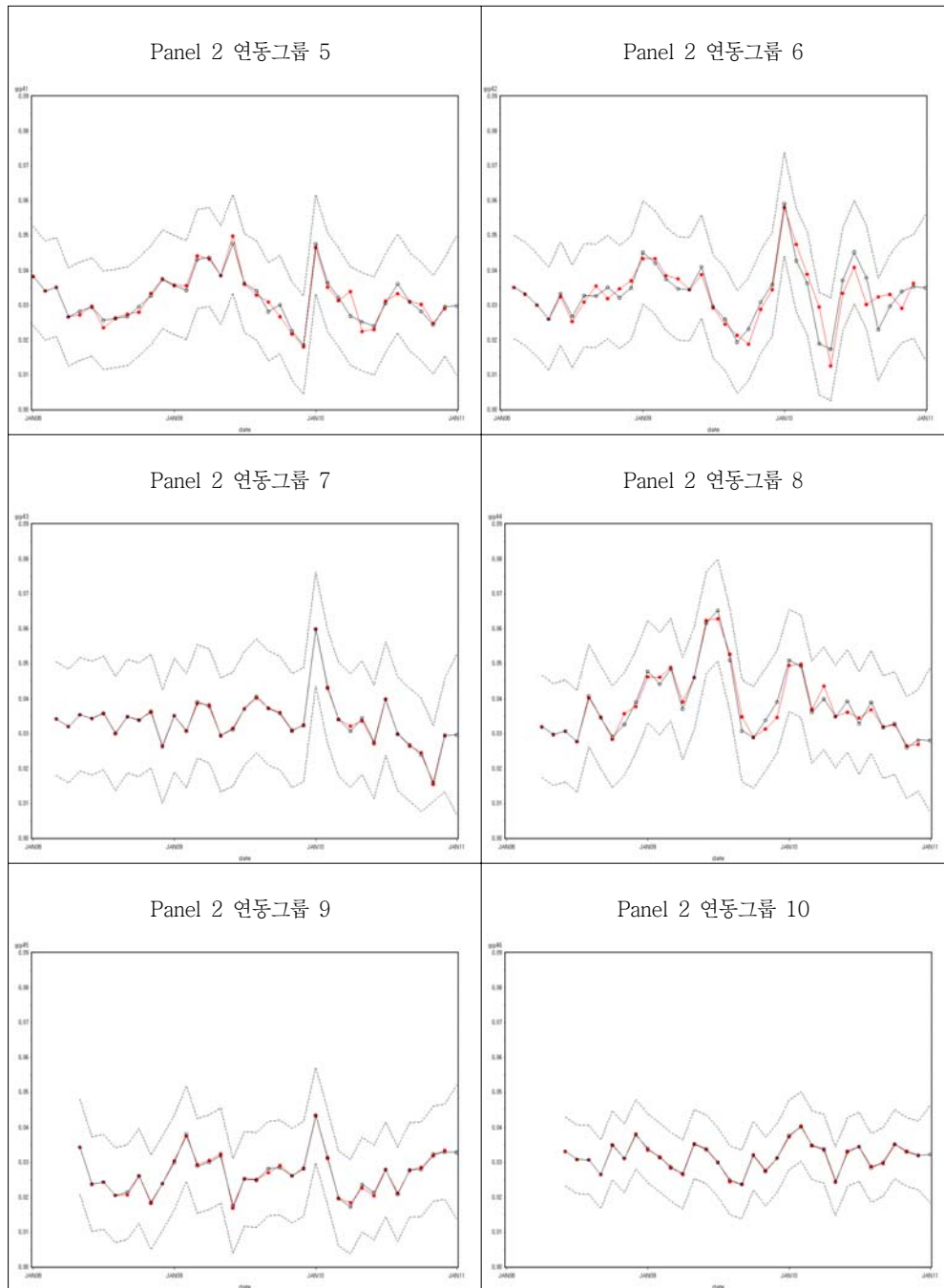
[그림 4-8]은 24개월 이상 응답한 연동그룹 24개에서 산출된 실업률 자료와 ARIMA(4,1,0) 모형을 이용하여 적합 결과의 예측값과 이의 신뢰구간을 나타내고 있다. 모든 연동그룹에서 시계열 모형의 적합이 잘 이루어진 것으로 판단된다. 간단한 시계열 모형 적합을 통해서 볼 때 시계열 모형을 이용한 분산-공분산 행렬의 추정 및 최량선형불편추정량의 사용 역시 가능한 하나의 방안이 될 수 있으리라 여겨진다. 그러나 이를 위해서는 각 관심모수 별 모형적합, 모형 타당성 검토 등의 복잡한 절차가 필요하며 특별히 모형 기반 추정량들의 평가가 장기간 이루어진 후에 적용되는 것이 바람직하리라 여겨진다.



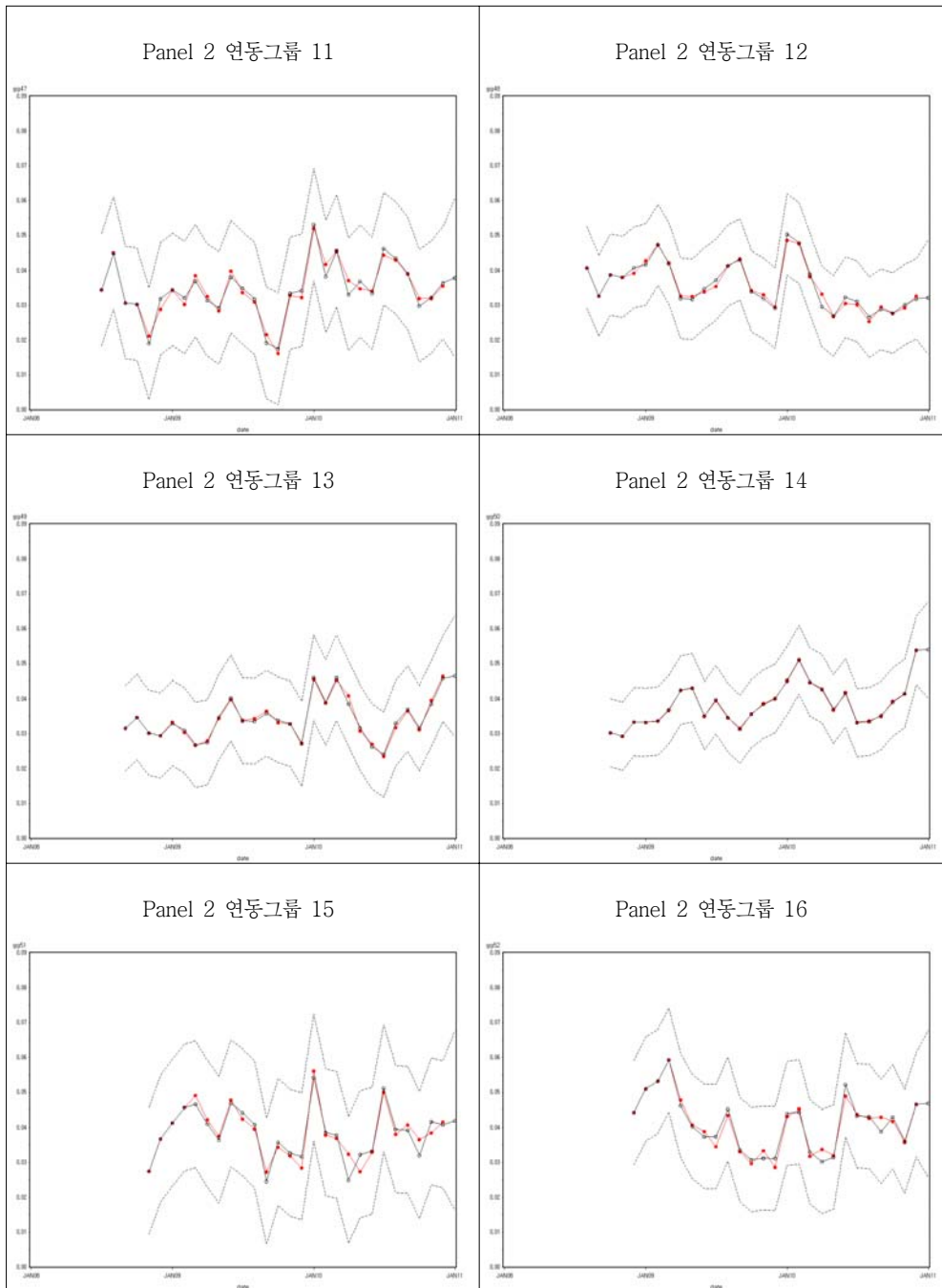
[그림 4-8] 각 연동그룹 추정량들의 ARI(4,1)모형 적합 결과



[그림 4-8] 각 연동그룹 추정량들의 ARI(4,1)모형 적합 결과 (계속)



[그림 4-8] 각 연동그룹 추정량들의 ARI(4,1)모형 적합 결과 (계속)



[그림 4-8] 각 연동그룹 추정량들의 ARI(4,1)모형 적합 결과 (계속)

4. 분산추정식

1) 개론

조사된 자료를 이용하여 추정량의 분산을 추정하기 위한 분산 추정식은 일반적으로 표본 추출 시 사용된 표본추출방법과 모수를 추정하기 위한 추정식에 의해 결정된다. 경제활동인구조사처럼 복합표본설계(complex sample design)를 이용하여 표본이 추출되고, 비선형(non-linear) 추정방법인 사후층화추정식(post-stratified estimator)을 사용하여 모수가 추정된 경우 계산 가능한 불편 분산추정식의 산출이 사실상 불가능하다. 이러한 경우, 분산을 추정하기 위한 방법으로 테일러 전개(Taylor expansion)에 근거한 선형근사(linear approximation) 분산 추정방법과 추출된 표본으로부터 부분 표본을 반복적으로 추출하고 각 부분 표본을 사용한 추정량의 변이를 측정하여 분산을 추정하는 반복 분산추정방법(replicate variance estimation) 중 잭나이프 분산추정방법(Jackknife variance estimation)을 고려할 수 있다.

본 절에서는 경제활동인구조사의 관심모수 추정량으로 사용된 사후층화추정량의 분산추정량으로 고려될 수 있는 선형근사 분산추정법과 잭나이프 분산추정방법을 단순한 표본추출방법인 층화단순임의추출법(stratified simple random sampling)하에서 설명하고 현행 통계청에서 가정하고 있는 균등확률표본(equal probability sample)을 만족하는 층화집락추출법(stratified cluster sampling) 하에서의 선형근사 분산추정방법과 잭나이프 분산추정방법을 제안하고 이의 통계적 성질들을 분석하였다. 특별히 현행 통계청에서 사용하고 있는 잭나이프 분산추정방안이 가정하고 있는 표본설계 하에서도 이론적으로 타당하지 않음을 보이고 복잡한 형태의 반복 추정방안인 잭나이프 분산추정방안 대신 사용가능한 선형근사 분산추정방안을 본 절에서는 제시하려 한다. 이를 위하여 경제활동인구조사 결과 자료 분석 및 모의실험을 수행하였다.

더불어 현행 통계청 경제활동인구조사의 표본 설계에 대한 가정이 타당하지 않는 경우, 즉 경제활동인구조사 표본이 인구주택총조사의 10% 자료로부터 추출된 이상 표본(two phase sample)이며 또한 균등확률표본이 아닌 경우 사용할 수 있는 잭나이프 분산 추정량의 일반적인 형태를 본 절에서는 제시하려 한다.

2) 층화단순임의추출 하에서의 분산추정

분산추정방안에 대한 설명을 위해서 먼저 본 연구에서 사용할 추정량이 사후층화추정량임을 감안하여 모수 추정식 및 분산추정식을 위한 수학적 기호들을 먼저 정의한다. 층화단순임의추출을 위하여 사용된 층을 나타내는 표본설계 층의 기호로는 $h = 1, \dots, H$ 를 사용한다. 즉 표본설계를 위하여서는 전체 H 개의 층을 사용한다고 가정한다. 또한 추정량의 정의 시 사용하는 사후 층을 나타내기 위해서는 $l = 1, \dots, L$ 을 사용한다. 따라서 주어진 모집단 U 와 표본 S 의 표본설계 층과 사후 층을 나타내기 위한 구조적인 모습을 그림으로 표현하면 다음과 같다. 표본크기와 모집단의 크기를 나타내기 위한 부호로는 각각 n 과 N 을 사용한다.

		사후 층				
		1	...	l	...	L
표 본 설 계 층	1	$\ddots$				$n_{1.}$ $N_{1.}$
	$\vdots$					
	h			N_{hl} n_{hl}		$\vdots$ $\vdots$
	$\vdots$					
	H				$\ddots$	$n_{H.}$ $N_{H.}$
		$n_{.1}$	...		$n_{.L}$	$n_{..}$ (= n)
		$N_{.1}$	...		$N_{.L}$	$N_{..}$ (= N)

[그림 4-9] 모집단 구조

층화단순임의표본과 주어진 사후 층의 구조를 이용하여 정의되는 사후층화추정량은 다음과 같다.

$$\bar{y}_{post} = \sum_{i \in S} w_i y_i = \sum_{l=1}^L \frac{N_{.l}}{N} \tilde{y}_{.l},$$

위 식에서 h 번째 표본설계 층 및 l 번째 사후 층에 속한 관측치에 부여되는 가중치 w_i 는

$$\begin{aligned} w_i &= \frac{1}{N} w_{0i} g_{is} \\ &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} \left(1 + \left(\underline{t}_x - \hat{\underline{t}}_{x\pi} \right) (X' W_0 X)^{-1} \underline{x}_i' \right) \\ &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} \left(1 + \left((N_{.1}, \dots, N_{.L}) - (\tilde{n}_{.1}, \dots, \tilde{n}_{.L}) \right) \begin{pmatrix} \tilde{n}_{.1} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \tilde{n}_{.l} & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & 0 & \tilde{n}_{.L} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \\ &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} + \frac{1}{N} \frac{N_{h.}}{n_{h.}} (N_{.l} - \tilde{n}_{.l}) (\tilde{n}_{.l})^{-1} \\ &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} + \frac{1}{N} \frac{N_{h.}}{n_{h.}} N_{.l} (\tilde{n}_{.l})^{-1} - \frac{1}{N} \frac{N_{h.}}{n_{h.}} \\ &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} \frac{N_{.l}}{\tilde{n}_{.l}} \end{aligned}$$

이다. 여기서 사용된 각각의 수학적 표현은

$$\begin{aligned} \tilde{y}_{.l} &= \frac{1}{\tilde{n}_{.l}} \sum_{h=1}^H \frac{N_{h.}}{n_{h.}} n_{hl} \bar{y}_{hl}, \\ g_{is} &= 1 + \left(\underline{t}_x - \hat{\underline{t}}_{x\pi} \right) (X' W_0 X)^{-1} \underline{x}_i', \\ \underline{x}_i' &= (\delta_{i1}, \dots, \delta_{il}, \dots, \delta_{iL})', \\ \delta_{il} &= \begin{cases} 1, & i \in S_l \\ 0, & i \notin S_l \end{cases} \\ \tilde{n}_{.l} &= \sum_{h=1}^H n_{hl} \frac{N_{h.}}{n_{h.}}, \end{aligned}$$

w_{0i} 는 표본설계 가중치 $N_{h.}/n_{h.}$ 을, $\bar{y}_{hl}$ 는 h 번째 표본설계 층 및 l 번째 사후 층에 속한 관측치들을 이용하여 계산된 단순 표본평균을 나타낸다.

층화단순임의추출 하에서 층별 표본배분방법으로 비례배분을 사용하는 경우, 즉

$N_{h.}/n_{h.} = N/n$ 일 때 사후층화추정량은 일반적으로 알려진 다음의 단순한 형태로 표현되며

$$\bar{y}_{post} = \sum_{i \in S} w_i y_i = \sum_{l=1}^L \frac{N_{.l}}{N} \tilde{y}_{.l},$$

여기서

$$w_i = \frac{N_{h.}}{n_{h.}} \frac{N_{.l}}{N} (\tilde{n}_{.l})^{-1} = \frac{N_{h.}}{n_{h.}} \frac{N_{.l}}{N} \left(\sum_{h=1}^H n_{hl} \frac{N_h}{n_h} \right)^{-1} = \frac{N_{.l}}{N} \frac{1}{n_{.l}},$$

$$\tilde{y}_{.l} = \frac{1}{n_{.l}} \sum_{h=1}^H n_{hl} \bar{y}_{hl}$$

을 각 각 나타낸다.

사후층화추정량은 기본적으로 비선형 추정량이며 따라서 일반적으로 사용되는 분산식, 특별히 SAS의 `Surveymeans` 프로시저를 이용하여 계산되어지는 분산추정량은 실제 분산을 과대 추정한다. 층화단순임의추출 하에서 정의된 사후층화추정량의 선형근사 분산추정량으로는 다음을 고려할 수 있다. 주어진 분산추정량은 Samdal, Swensson & Wretman (1992)의 6장에 자세히 설명되어 있다. 또한 이 분산추정량은 SAS의 `Surveyreg` 프로시저를 이용하여 계산되어지며 `Surveyreg` 프로시저의 사용을 위해서는 위에서 정의된 각 관측치별 벡터 $\underline{x}$ 을 지정해주어야 한다.

$$\hat{V}_{Taylor}(\bar{y}_{post}) = \frac{1}{N^2} \sum_{h=1}^H N_h^2 \left(1 - \frac{n_h}{N_h} \right) \frac{s_{ge,h}^2}{n_h},$$

여기서

$$s_{ge,h}^2 = \sum_{i \in S_h} (ge_i - \overline{ge}_h)^2 / (n_h - 1),$$

$$ge_i = g_{is} \times e_{is},$$

$$\overline{ge}_h = \frac{1}{n_h} \sum_{i \in S_h} ge_{is},$$

$$e_{is} = (y_i - \underline{x}_i' \hat{B}_0),$$

$$\hat{B}_0 = \left(\sum_{h=1}^H \sum_{i \in S_h} w_{0i} X_i' X_i \right)^{-1} \left(\sum_{h=1}^H \sum_{i \in S_h} w_{0i} X_i' y_i \right),$$

잭나이프 분산추정은 표본에 주어진 1차 추출단위(Primary Sampling Unit)을 한 개씩 삭제한 후 모수의 추정량을 산출하고 이렇게 산출된 n 개의 추정량의 변이를 이용하여 분산을 추정하는 방안이다. 또한 잭나이프 분산추정량은 각 1차 추출단위가 제거되었을 때 남아있는 $n-1$ 개의 1차 추출단위에 부여될 가중치인 잭나이프 반복 가중치(Jackknife replicate weight)를 산출함으로 모든 관심 모수에 대한 추정량

의 분산추정량을 구할 수 있는 장점이 있어 현실적으로 많이 사용되고 있다. 고려된 사후층화추정량의 층화단순임의추출법 하에서의 잭나이프 분산추정량은 다음과 같이 정의될 수 있다.

$$\hat{V}_{JK}(\bar{y}_{post}) = \sum_{h=1}^H \frac{n_h - 1}{n_h} \sum_{k \in S_h} \left(\bar{y}_{post}^{(-k)} - \bar{y}_{post} \right)^2,$$

여기서

$$\bar{y}_{post} = \frac{1}{n} \sum_{k=1}^n \bar{y}_{post}^{(-k)},$$

$$\bar{y}_{post}^{(-k)} = \sum_{h=1}^H \sum_{\substack{i \in S_h \\ i \neq k}} w_i^{(-k)} y_i,$$

그리고 $w_i^{(-k)}$ 는 표본 층 h' 과 사후 층 l' 에 있는 k 번째 1차 추출단위를 삭제 후 사후층화추정량 산출을 위해 i 번째 관측치에 부여되는 잭나이프 반복 가중치이다. 본 절에서 고려하고 있는 층화단순임의추출법 하에서의 1차 추출단위는 관측단위와 일치하며 이 때 정의되는 잭나이프 반복 가중치의 일반적인 표현은 다음과 같다.

$$\begin{aligned} w_i^{(-k)} &= \frac{1}{N} w_{0i}^{(-k)} g_{is}^{(-k)} \\ &= \frac{1}{N} w_{0i}^{(-k)} \left(1 + \left(\bar{t}_x - \hat{\bar{t}}_{x\pi}^{(-k)} \right) \left(X'^{(-k)} W_0^{(-k)} X^{(-k)} \right)^{-1} \bar{x}_i' \right) \\ &= \frac{1}{N} w_{0i}^{(-k)} + w_{0i}^{(-k)} \left(\bar{x}_N - \hat{\bar{x}}_{st}^{(-k)} \right) \left(X'^{(-k)} W_0^{(-k)} X^{(-k)} \right)^{-1} \bar{x}_i', \end{aligned}$$

여기서

$$w_{0i}^{(-k)} = \begin{cases} \frac{N_{h.}}{n_{h.}}, & i \notin S_{h'} \\ \frac{N_{h'}}{(n_{h'} - 1)}, & i \in S_{h'} \end{cases},$$

$$\begin{aligned} \hat{\bar{x}}_{st}^{(-k)} &= \sum_{\substack{h=1 \\ h \neq h'}}^H \sum_{i \in S_h} \frac{1}{N} \frac{N_{h.}}{n_{h.}} \bar{x}_i + \frac{1}{N} \frac{N_{h'}}{(n_{h'} - 1)} \sum_{\substack{i \in S_{h'} \\ i \neq k}} \bar{x}_i \\ &= \sum_{\substack{h=1 \\ h \neq h'}}^H \frac{1}{N} \frac{N_{h.}}{n_{h.}} (n_{h1}, \dots, n_{hL}) + \frac{1}{N} \frac{N_{h'}}{(n_{h'} - 1)} (n_{h'1}, \dots, n_{h'l'} - 1, \dots, n_{h'L}), \\ X'^{(-k)} W_0^{(-k)} X^{(-k)} &= \begin{pmatrix} \bar{x}'_1, \dots, \bar{x}'_{h'}, \dots, \bar{x}'_H \end{pmatrix} \begin{pmatrix} \frac{N_{h.}}{n_{h.}} I_{(n-1) \times (n-1)} \end{pmatrix} \\ &\quad \begin{pmatrix} \bar{x}'_1, \dots, \bar{x}'_{h'}, \dots, \bar{x}'_H \end{pmatrix}' \end{aligned}$$

$$\begin{aligned}
&= \sum_{\substack{h=1 \\ h \neq h'}}^H \underline{x}_{h.}' W_{0,h.} \underline{x}_{h.} + \underline{x}_{h'.}^{(-k)'} W_{0,h'.}^{(-k)} \underline{x}_{h'.}^{(-k)} \\
&= \sum_{\substack{h=1 \\ h \neq h'}}^H \frac{N_{h.}}{n_{h.}} \begin{pmatrix} n_{h1} & & 0 \\ & \ddots & \\ 0 & & n_{hL} \end{pmatrix} + \frac{N_{h'.}}{(n_{h'.} - 1)} \begin{pmatrix} n_{h'1} & & & 0 \\ & \ddots & & \\ & & n_{h'l'} - 1 & \\ 0 & & & n_{h'L} \end{pmatrix} \\
&= \begin{pmatrix} \hat{t}_{x\pi,1} & & & 0 \\ & \ddots & & \\ & & \hat{t}_{x\pi,l} & \\ & & & \ddots \\ 0 & & & & \hat{t}_{x\pi,L} \end{pmatrix}, \\
&\begin{cases} \hat{t}_{x\pi,l} = \sum_{\substack{h=1 \\ h \neq h'}}^H \frac{N_{h.}}{n_{h.}} n_{hl} + \frac{N_{h'.}}{(n_{h'.} - 1)} n_{h'l}, & k \notin PS_{l'} \\ \hat{t}_{x\pi,l'} = \sum_{\substack{h=1 \\ h \neq h'}}^H \frac{N_{h.}}{n_{h.}} n_{hl'} + \frac{N_{h'.}}{(n_{h'.} - 1)} (n_{h'l'} - 1), & k \in PS_{l'} \end{cases}, \\
\underline{x}_N - \widehat{\underline{x}}_{st}^{(-k)} = \begin{cases} a_l = \frac{N_{.l}}{N} - \frac{1}{N} \hat{t}_{x\pi,l}, & k \notin PS_{l'} \\ a_{l'} = \frac{N_{.l'}}{N} - \frac{1}{N} \hat{t}_{x\pi,l'}, & k \in PS_{l'} \end{cases}.
\end{aligned}$$

표본 중 h' 와 사후 중 l' 에 있는 k 번째 단위를 삭제하는 경우 i 번째 관측 단위
에 부여되는 잼나이프 반복 가중치는 각 관측치의 표본설계 중 $S_{h'}$ 과 사후 중 $PS_{l'}$
에 포함여부 따라 4가지의 경우로 분류되고 위의 주어진 잼나이프 반복 가중치를
각 경우 별로 계산하면 다음과 같다.

(1) $i \notin S_{h'}$, $i \notin PS_{l'}$

$$\begin{aligned}
w_i^{(-k)} &= \frac{1}{N} w_{0i}^{(-k)} g_{is}^{(-k)} \\
&= \frac{1}{N} w_{0i}^{(-k)} + w_{0i}^{(-k)} \left(\underline{x}_N - \widehat{\underline{x}}_{st}^{(-k)} \right) \left(X'^{(-k)} W_0^{(-k)} X^{(-k)} \right)^{-1} \underline{x}_i' \\
&= \frac{1}{N} \frac{N_{h.}}{n_{h.}} + \frac{N_{h.}}{n_{h.}} (a_1, \dots, a_l, \dots, a_L) \begin{pmatrix} \hat{t}_{x\pi,1} & & & 0 \\ & \ddots & & \\ & & \hat{t}_{x\pi,l} & \\ & & & \ddots \\ 0 & & & & \hat{t}_{x\pi,L} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} + \frac{N_{h.}}{n_{h.}} a_l \hat{t}_{x\pi, l}^{-1} \\
 &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} + \frac{N_{h.}}{n_{h.}} \left(\frac{N_{.l}}{N} - \frac{1}{N} \hat{t}_{x\pi, l} \right) \hat{t}_{x\pi, l}^{-1} \\
 &= \frac{N_{h.}}{n_{h.}} \frac{N_{.l}}{N} \hat{t}_{x\pi, l}^{-1} \\
 &= \frac{N_{.l}}{\sum_{\substack{h=1 \\ h \neq h'}}^H \frac{N_{h.}}{n_{h.}} n_{hl} + \frac{N_{h'.}}{(n_{h'.} - 1) n_{h'l}}} \frac{N_{h.}}{n_{h.}} \frac{1}{N}.
 \end{aligned}$$

(2) $i \notin S_{h'}, i \in PS_{l'}$

$$\begin{aligned}
 w_i^{(-k)} &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} + \frac{N_{h.}}{n_{h.}} a_{l'} \hat{t}_{x\pi, l'}^{-1} \\
 &= \frac{1}{N} \frac{N_{h.}}{n_{h.}} + \frac{N_{h.}}{n_{h.}} \left(\frac{N_{.l'}}{N} - \frac{1}{N} \hat{t}_{x\pi, l'} \right) \hat{t}_{x\pi, l'}^{-1} \\
 &= \frac{N_{h.}}{n_{h.}} \frac{N_{.l'}}{N} \hat{t}_{x\pi, l'}^{-1} \\
 &= \frac{N_{.l'}}{\sum_{\substack{h=1 \\ h \neq h'}}^H \frac{N_{h.}}{n_{h.}} n_{hl'} + \frac{N_{h'.}}{(n_{h'.} - 1) (n_{h'l'} - 1)}} \frac{N_{h.}}{n_{h.}} \frac{1}{N}.
 \end{aligned}$$

(3) $i \in S_{h'}, i \notin PS_{l'}$

$$\begin{aligned}
 w_i^{(-k)} &= \frac{1}{N} \frac{N_{h'.}}{(n_{h'.} - 1)} + \frac{N_{h'.}}{(n_{h'.} - 1)} a_l \hat{t}_{x\pi, l}^{-1} \\
 &= \frac{1}{N} \frac{N_{h'.}}{(n_{h'.} - 1)} + \frac{N_{h'.}}{(n_{h'.} - 1)} \left(\frac{N_{.l}}{N} - \frac{1}{N} \hat{t}_{x\pi, l} \right) \hat{t}_{x\pi, l}^{-1} \\
 &= \frac{N_{h'.}}{(n_{h'.} - 1)} \frac{N_{.l}}{N} \hat{t}_{x\pi, l}^{-1} \\
 &= \frac{N_{.l}}{\sum_{\substack{h=1 \\ h \neq h'}}^H \frac{N_{h.}}{n_{h.}} n_{hl} + \frac{N_{h'.}}{(n_{h'.} - 1) n_{h'l}}} \frac{N_{h'.}}{(n_{h'.} - 1)} \frac{1}{N}.
 \end{aligned}$$

(4) $i \in S_{h'}$, $i \in PS_{l'}$

$$\begin{aligned}
 w_i^{(-k)} &= \frac{1}{N} \frac{N_{h'}}{(n_{h'} - 1)} + \frac{N_{h'}}{(n_{h'} - 1)} a_{l'} \hat{t}_{x\pi, l'}^{-1} \\
 &= \frac{1}{N} \frac{N_{h'}}{(n_{h'} - 1)} + \frac{N_{h'}}{(n_{h'} - 1)} \left(\frac{N_{l'}}{N} - \frac{1}{N} \hat{t}_{x\pi, l'} \right) \hat{t}_{x\pi, l'}^{-1} \\
 &= \frac{N_{h'}}{(n_{h'} - 1)} \frac{N_{l'}}{N} \hat{t}_{x\pi, l'}^{-1} \\
 &= \frac{N_{l'}}{\sum_{\substack{h=1 \\ h \neq h'}}^H \frac{N_h}{n_h} n_{hl'} + \frac{N_{h'}}{(n_{h'} - 1)} (n_{h'l'} - 1)} \frac{N_{h'}}{(n_{h'} - 1)} \frac{1}{N}.
 \end{aligned}$$

3) 총화집락추출 하에서의 분산추정

2절에서 총화단순임의추출법 하에서의 사후총화추정량의 가능한 분산추정량을 고려하였다. 본 절에서는 이의 결과를 확대하여 현행 경제활동인구조사에서 가정하고 있는 총화집락추출법 하에서의 분산추정방안을 고려한다. 현행 경제활동인구조사의 표본 추출을 위하여서는 2상 표본추출방법이 사용되었으나 1상과 2상에서의 표본설계 총과 표본추출단위가 동일하다는 가정 하에서 사후총화추정량의 분산추정량을 고려하였다. 본 절에서는 통계청의 분산추정을 위한 가정이 적절하다는 조건 하에서 사용 가능한 분산추정량으로 앞 절에서와 마찬가지로 선형근사 분산추정량과 잭나이프 분산추정량을 고려한다. 먼저 각 가구원에 부여되는 표본설계 가중치를 살펴보면 다음과 같다.

$$w_{1i} = \frac{\sum_{k=1}^{N_h} z_{hk}}{n_h z_{hk}} \times \frac{z_{hk}}{20} = \frac{\sum_{k=1}^{N_h} z_{hk}}{n_h 20} = \frac{Z_h}{n_h 20},$$

여기서 N_h 는 h 번째 층의 모집단 조사구 수를, n_h 는 h 번째 층의 표본 조사구 수를, z_{hk} 는 h 번째 층의 k 번째 조사구의 가구 크기를 그리고 Z_h 는 h 번째 층의 가구 크기를 각 각 나타낸다. 이는 각 층별 조사구 추출이 조사구 내의 가구 수에 비례한 확률비례추출법에 의하여 추출되었고 각 조사구 내의 20개 가구는 균등 확률을 갖는 일종의 단순임의추출법을 통해 추출되었기 때문이다. 주어진 표본 추출방안 하에서 계산되어진 동일 층 내의 모든 가구원 표본 가중치는 그 표본설계의 특징으로 인하여 동일한 값을 갖게 된다.

주어진 표본설계 하에서의 사후총화추정량은 다음과 같이 정의된다.

$$\bar{y}_{post} = \sum_{i \in S} w_i y_i = \sum_{l=1}^L \frac{M_{.l}}{M} \tilde{y}_{.l}$$

사후총화추정량을 정의하기 위하여 정의된 h 번째 설계 층 및 l 번째 사후 층에 속한 관측치 i 에 부여되는 가중치는

$$\begin{aligned} w_i &= \frac{1}{M} w_{1i} g_{is} \\ &= \frac{1}{M} \frac{Z_h}{n_h 20} \left(1 + \left(\underline{t}_x - \hat{t}_{x\pi} \right) (X' W_1 X)^{-1} \underline{x}_i' \right) \\ &= \frac{1}{M} \frac{Z_h}{n_h 20} \left(1 + \left((M_{.1}, \dots, M_{.L}) - (\tilde{m}_{.1}, \dots, \tilde{m}_{.L}) \right) \begin{pmatrix} \tilde{m}_{.1} & 0 & \dots & \dots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & \tilde{m}_{.l} & & \vdots \\ \vdots & & & \ddots & 0 \\ 0 & \dots & \dots & 0 & \tilde{m}_{.L} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) \\ &= \frac{1}{M} \frac{Z_h}{n_h 20} + \frac{1}{M} \frac{Z_h}{n_h 20} (M_{.l} - \tilde{m}_{.l}) (\tilde{m}_{.l})^{-1} \\ &= \frac{1}{M} \frac{Z_h}{n_h 20} \frac{M_{.l}}{\tilde{m}_{.l}}, \end{aligned}$$

로써 여기서

$$\begin{aligned} g_{is} &= \frac{Z_h}{n_h 20} \left(1 + \left(\underline{t}_x - \hat{t}_{x\pi} \right) (X' W_1 X)^{-1} \underline{x}_i' \right) \\ \tilde{y}_{.l} &= \frac{1}{\tilde{m}_{.l}} \sum_{h=1}^H \frac{Z_h}{n_h 20} m_{hl} \bar{y}_{hl}, \\ \tilde{m}_{.l} &= \sum_{h=1}^H m_{hl} \frac{Z_h}{n_h 20}, \end{aligned}$$

M 는 모집단 경제활동인구 수를, $M_{.l}$ 은 l 번째 사후 층의 모집단 경제활동인구 수를 그리고 m_{hl} 은 h 번째 층의 l 번째 사후 층의 표본 경제활동인구 수를 나타낸다.

총화집락추출표본을 이용하여 정의된 사후총화추정량의 분산추정량으로는 앞에서와 마찬가지로 선형근사 분산추정량과 잭나이프 분산추정량을 고려할 수 있다. 먼저 선형근사 추정량의 정의를 위하여 위에서 정의된 사후총화추정량을 다음의 단순한 형태의 회귀추정량으로 표현하였다.

$$\bar{y}_{post} = \underline{x}_N' \hat{\underline{\beta}}.$$

여기서 $\hat{\underline{\beta}} = (X' W_1 X)^{-1} (X' W_1 y)$ 이며, $\underline{x}_N$ 는 $\underline{x}_i$ 의 모집단 평균을 나타내고 W_1 은 w_{1i} 를 대각원소로 갖는 대각행렬을 나타낸다. 따라서 사후총화추정량의 선형근사 분

산추정량은 다음과 같이 정의된다.

$$\hat{V}_{Taylor}(\bar{y}_{post}) = \bar{x}_N' \hat{V}(\hat{\underline{\beta}}) \bar{x}_N.$$

회귀계수 추정량 $\hat{\underline{\beta}}$ 의 분산추정량은

$$\hat{V}(\hat{\underline{\beta}}) = (X' W_1 X)^{-1} G (X' W_1 X)^{-1},$$

으로 여기서

$$G = \sum_{h=1}^H \frac{n_h}{n_{h-1}} \sum_{i=1}^{n_h} (\underline{e}_{hi.} - \bar{e}_{h..})' (\underline{e}_{hi.} - \bar{e}_{h..})$$

이며 h 번째 설계 층의 i 번째 집락(조사구)내 j 번째 가구에 정의되는 통계량들은 다음과 같다.

$$\underline{e}_{hij} = w_{1,hij} r_{hij} \underline{x}_{hij},$$

$$\underline{e}_{hi.} = \sum_{j=1}^{m_{hi}} \underline{e}_{hij},$$

$$\bar{e}_{h..} = \frac{1}{n_h} \sum_{i=1}^{n_h} \underline{e}_{hi.},$$

$$\underline{r} = (r_{111}, \dots, r_{H n_h m_{H n_h}})' = \underline{y} - X \hat{\underline{\beta}}.$$

위의 선형근사 분산추정량은 층화단순임의추출의 경우와 마찬가지로 사후 층을 나타내는 지시함수로 구성된 $\underline{x}$ 벡터를 정의하고 이를 SAS surveyreg 프로시저에 적용하여 산출할 수 있다.

잭나이프 분산추정량을 구하기 위하여서는 각 1차 추출단위인 조사구를 삭제한 후 잭나이프 반복 가중치를 산출하고 이를 이용하여 계산되어진 표본 조사구 수와 동일한 사후층화추정량들의 변동을 계산하여야 한다. 이러한 과정을 통해 정의되는 잭나이프 분산추정량은 다음과 같다.

$$\hat{V}_{JK}(\bar{y}_{post}) = \sum_{h=1}^H \frac{n_h - 1}{n_h} \sum_{k \in S_h} (\bar{y}_{post}^{(-k)} - \bar{y}_{post})^2,$$

여기서

$$\bar{y}_{post} = \frac{1}{n} \sum_{k=1}^n \bar{y}_{post}^{(-k)},$$

$$\bar{y}_{post}^{(-k)} = \sum_{h=1}^H \sum_{\substack{i \in S_h \\ i \neq k}} w_i^{(-k)} y_i,$$

이며 n_h 는 h 번째 설계 층의 조사구 수를 나타내며 $w_i^{(-k)}$ 는 k 번째 단위를 삭제 후 계산된 표본설계 층 h' 에 속한 i 번째 단위의 잭나이프 반복 가중치로서 다음과

같다.

$$\begin{aligned}
 w_i^{(-k)} &= \frac{1}{M} w_{1i}^{(-k)} g_{is}^{(-k)} \\
 &= \frac{1}{M} w_{1i}^{(-k)} \left(1 + \left(\underline{t}_x - \hat{t}_{x\pi}^{(-k)} \right) \left(X'^{(-k)} W_1^{(-k)} X^{(-k)} \right)^{-1} \underline{x}_i' \right) \\
 &= \frac{1}{M} w_{1i}^{(-k)} + w_{1i}^{(-k)} \left(\underline{x}_M - \widehat{\underline{x}}_{st}^{(-k)} \right) \left(X'^{(-k)} W_1^{(-k)} X^{(-k)} \right)^{-1} \underline{x}_i',
 \end{aligned}$$

여기서

$$\begin{aligned}
 w_{1i}^{(-k)} &= \begin{cases} \frac{Z_{h.}}{n_{h.}20}, & i \notin S_{h'} \\ \frac{Z_{h'}}{(n_{h'} - 1)20}, & i \in S_{h'} \end{cases}, \\
 \widehat{\underline{x}}_{st}^{(-k)} &= \sum_{\substack{h=1 \\ h \neq h'}}^H \sum_{i \in S_h} \frac{1}{M} \frac{Z_{h.}}{n_{h.}20} x_i + \frac{1}{M} \frac{Z_{h'}}{(n_{h'} - 1)20} \sum_{\substack{i \in S_{h'} \\ i \neq k}} x_i \\
 &= \sum_{\substack{h=1 \\ h \neq h'}}^H \frac{1}{M} \frac{Z_{h.}}{n_{h.}20} (m_{h1}, \dots, m_{hL}) \\
 &\quad + \frac{1}{M} \frac{Z_{h'}}{(n_{h'} - 1)20} (m_{h'1} - m_{h'j1}, \dots, m_{h'l'} - m_{h'jl'}, \dots, m_{h'L} - m_{h'jL}), \\
 X'^{(-k)} W_1^{(-k)} X^{(-k)} &= \sum_{\substack{h=1 \\ h \neq h'}}^H \underline{x}_{h.}' W_{1,h.} \underline{x}_{h.} + \underline{x}_{h'}'^{(-k)} W_{1,h'}^{(-k)} \underline{x}_{h'}^{(-k)} \\
 &= \sum_{\substack{h=1 \\ h \neq h'}}^H \frac{Z_{h.}}{n_{h.}20} \begin{pmatrix} m_{h1} & & 0 \\ & \ddots & \\ 0 & & m_{hL} \end{pmatrix} \\
 &\quad + \frac{Z_{h'}}{(n_{h'} - 1)20} \begin{pmatrix} m_{h'1} - m_{h'j1} & & & & 0 \\ & \ddots & & & \\ & & m_{h'l'} - m_{h'jl'} & & \\ & & & \ddots & \\ 0 & & & & m_{h'L} - m_{h'jL} \end{pmatrix} \\
 &= \begin{pmatrix} \hat{t}_{x,1} & & & & 0 \\ & \ddots & & & \\ & & \hat{t}_{x,l} & & \\ 0 & & & \ddots & \\ & & & & \hat{t}_{x,L} \end{pmatrix},
 \end{aligned}$$

$$\begin{cases} \hat{t}_{x,l} = \sum_{\substack{h=1 \\ h \neq h'}}^H \frac{Z_{h.}}{n_{h.}20} m_{hl} + \frac{Z_{h'.}}{(n_{h'.}-1)20} m_{h'l}, & k \notin PS_{l'} \\ \hat{t}_{x,l'} = \sum_{\substack{h=1 \\ h \neq h'}}^H \frac{Z_{h.}}{n_{h.}20} m_{hl'} + \frac{Z_{h'.}}{(n_{h'.}-1)20} (m_{h'l'} - m_{h'jl'}), & k \in PS_{l'} \end{cases},$$

$$\bar{x}_M - \bar{x}_{st}^{(-k)} = \begin{cases} a_l = \frac{M_{.l}}{M} - \frac{1}{M} \hat{t}_{x,l}, & k \notin PS_{l'} \\ a_{l'} = \frac{M_{.l'}}{M} - \frac{1}{M} \hat{t}_{x,l'}, & k \in PS_{l'} \end{cases},$$

그리고 $m_{h'jl'}$ 는 h' 번째 표본설계 층 j 번째 가구의 l' 번째 사후 층에 속한 표본 경제활동인구 수를 나타낸다.

층화단순임의추출법의 경우와 달리 잭나이프 반복 가중치 산출을 위한 삭제단위 (조사구)가 사후 층 구분의 기본단위인 관측단위(가구원)와 다르기 때문에 단순한 형태의 반복 가중치 계산이 매우 복잡하다. 표본설계 층 $S_{h'}$ 에 속한 조사구 k 가 잭나이프 반복 가중치 산출을 위해 삭제되었을 때 각 개인이 부여받게 되는 잭나이프 반복 가중치는 삭제 조사구 내의 가구원들의 사후 층 분포에 따라 결정된다. k 번째 조사구가 삭제되었을 때 가능한 4가지 경우를 고려하여 잭나이프 반복 가중치를 단순화하면 다음과 같다.

- (1) $i \notin S_{h'}$, $i \notin PS_{l'}$ (i 번째 가구원이 k 번째 조사구가 삭제된 표본설계 층 $S_{h'}$ 에 속하지 않고 k 번째 조사구에 속한 가구원의 사후 층 $PS_{l'}$ 에도 속하지 않는 경우)

$$\begin{aligned} w_i^{(-k)} &= \frac{1}{M} w_{1i}^{(-k)} g_{is}^{(-k)} \\ &= \frac{1}{M} w_{1i}^{(-k)} + w_{1i}^{(-k)} \left(\bar{x}_M - \bar{x}_{st}^{(-k)} \right) \left(X'^{(-k)} W_1^{(-k)} X^{(-k)} \right)^{-1} \underline{x}_i' \\ &= \frac{1}{M} \frac{Z_{h.}}{n_{h.}20} + \frac{Z_{h.}}{n_{h.}20} (a_1, \dots, a_l, \dots, a_L) \begin{pmatrix} \hat{t}_{x,1} & & & 0 \\ & \ddots & & \\ & & \hat{t}_{x,l} & \\ 0 & & & \ddots & \hat{t}_{x,L} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &= \frac{1}{M} \frac{Z_{h.}}{n_{h.}20} + \frac{Z_{h.}}{n_{h.}20} a_l \hat{t}_{x,l}^{-1} \\ &= \frac{1}{M} \frac{Z_{h.}}{n_{h.}20} + \frac{Z_{h.}}{n_{h.}20} \left(\frac{M_{.l}}{M} - \frac{1}{M} \hat{t}_{x,l} \right) \hat{t}_{x,l}^{-1} \end{aligned}$$

$$\begin{aligned}
 &= \frac{Z_{h.}}{n_{h.}20} \frac{M_{.l}}{M} \hat{t}_{x,l}^{-1} \\
 &= \frac{\frac{M_{.l}}{\sum_{\substack{h=1 \\ h \neq h'}}^H \frac{Z_{h.}}{n_{h.}20} m_{hl} + \frac{Z_{h'.}}{(n_{h'.}-1)20} m_{h'l}}}{\frac{Z_{h.}}{n_{h.}20} m_{hl}} \frac{Z_{h.}}{n_{h.}20} \frac{1}{M} \\
 &= \frac{M_{.l}}{\frac{Z_{h.}}{n_{h.}20} m_{hl}} \frac{Z_{h.}}{n_{h.}20} \frac{1}{M} = \frac{M_{.l}}{m_{hl}} \frac{1}{M}.
 \end{aligned}$$

(2) $i \notin S_{h'}$, $i \in PS_{l'}$ (i 번째 가구원이 k 번째 조사구가 삭제된 표본설계 층 $S_{h'}$ 에 속하지 않는 반면 k 번째 조사구에 속한 가구원의 사후 층 $PS_{l'}$ 에는 속하는 경우)

$$\begin{aligned}
 w_i^{(-k)} &= \frac{1}{M} \frac{Z_{h.}}{n_{h.}20} + \frac{Z_{h.}}{n_{h.}20} a_{l'} \hat{t}_{x,l'}^{-1} \\
 &= \frac{1}{M} \frac{Z_{h.}}{n_{h.}20} + \frac{Z_{h.}}{n_{h.}20} \left(\frac{M_{.l'}}{M} - \frac{1}{M} \hat{t}_{x,l'} \right) \hat{t}_{x,l'}^{-1} \\
 &= \frac{Z_{h.}}{n_{h.}20} \frac{M_{.l'}}{M} \hat{t}_{x,l'}^{-1} \\
 &= \frac{\frac{M_{.l'}}{\sum_{\substack{h=1 \\ h \neq h'}}^H \frac{Z_{h.}}{n_{h.}20} m_{hl'} + \frac{Z_{h'.}}{(n_{h'.}-1)20} (m_{h'l'} - m_{h'kl'})}}{\frac{Z_{h.}}{n_{h.}20} m_{hl'}} \frac{Z_{h.}}{n_{h.}20} \frac{1}{M} \\
 &= \frac{M_{.l'}}{\frac{Z_{h.}}{n_{h.}20} m_{hl'}} \frac{Z_{h.}}{n_{h.}20} \frac{1}{M} = \frac{M_{.l'}}{m_{hl'}} \frac{1}{M}.
 \end{aligned}$$

(3) $i \in S_{h'}$, $i \notin PS_{l'}$ (i 번째 가구원이 k 번째 조사구가 삭제된 표본설계 층 $S_{h'}$ 에 속하고 k 번째 조사구에 속한 가구원의 사후 층 $PS_{l'}$ 에는 속하지 않는 경우)

$$\begin{aligned}
 w_i^{(-k)} &= \frac{1}{M} \frac{Z_{h'.}}{(n_{h'.}-1)20} + \frac{Z_{h'.}}{(n_{h'.}-1)20} a_l \hat{t}_{x,l}^{-1} \\
 &= \frac{1}{M} \frac{Z_{h'.}}{(n_{h'.}-1)20} + \frac{Z_{h'.}}{(n_{h'.}-1)20} \left(\frac{M_{.l}}{M} - \frac{1}{M} \hat{t}_{x,l} \right) \hat{t}_{x,l}^{-1}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{Z_{h'}.}{(n_{h'}. - 1)20} \frac{M_{.l}}{M} \hat{t}_{x,l}^{-1} \\
&= \frac{M_{.l}}{\sum_{\substack{h=1 \\ h \neq h'}}^H \frac{Z_{h.}}{n_{h.}20} m_{hl} + \frac{Z_{h'}.}{(n_{h'}. - 1)20} m_{h'l}} \frac{Z_{h'}.}{(n_{h'}. - 1)20} \frac{1}{M} \\
&= \frac{M_{.l}}{\frac{Z_{h'}.}{(n_{h'}. - 1)20} m_{h'l}} \frac{Z_{h'}.}{(n_{h'}. - 1)20} \frac{1}{M} = \frac{M_{.l}}{m_{h'l}} \frac{1}{M}.
\end{aligned}$$

(4) $i \in S_{h'}$, $i \in PS_{l'}$ (i 번째 가구원이 k 번째 조사구가 삭제된 표본설계 층 $S_{h'}$ 에 속하고 k 번째 조사구에 속한 가구원의 사후 층 $PS_{l'}$ 에도 속하는 경우)

$$\begin{aligned}
w_i^{(-k)} &= \frac{1}{M} \frac{Z_{h'}.}{(n_{h'}. - 1)20} + \frac{Z_{h'}.}{(n_{h'}. - 1)20} a_{l'} \hat{t}_{x,l'}^{-1} \\
&= \frac{1}{M} \frac{Z_{h'}.}{(n_{h'}. - 1)20} + \frac{Z_{h'}.}{(n_{h'}. - 1)20} \left(\frac{M_{.l'}}{M} - \frac{1}{M} \hat{t}_{x,l'} \right) \hat{t}_{x,l'}^{-1} \\
&= \frac{Z_{h'}.}{(n_{h'}. - 1)20} \frac{M_{.l'}}{M} \hat{t}_{x,l'}^{-1} \\
&= \frac{M_{.l'}}{\sum_{\substack{h=1 \\ h \neq h'}}^H \frac{Z_{h.}}{n_{h.}20} m_{hl'} + \frac{Z_{h'}.}{(n_{h'}. - 1)20} (m_{h'l'} - m_{h'kl'})} \frac{Z_{h'}.}{(n_{h'}. - 1)20} \frac{1}{M} \\
&= \frac{M_{.l'}}{\frac{Z_{h'}.}{(n_{h'}. - 1)20} (m_{h'l'} - m_{h'kl'})} \frac{Z_{h'}.}{(n_{h'}. - 1)20} \frac{1}{M} \\
&= \frac{M_{.l'}}{(m_{h'l'} - m_{h'kl'})} \frac{1}{M}.
\end{aligned}$$

경제활동인구조사 표본을 이용한 사후층회추정량의 분산추정량의 경우, 지역 변수가 표본 설계 층 및 사후 층으로 동시에 사용되었고 따라서 고려한 4가지의 경우 중 두 경우만이 실제 발생하고 두 경우에 해당하는 잭나이프 반복 가중치는 다음과 같이 표현된다.

(1) $i \in S_{h'}$, $i \in PS_{l'}$

$$w_i^{(-k)} = \frac{M_{.l'}}{(m_{h'l'} - m_{h'kl'})} \frac{1}{M}.$$

$$(2) i \notin S_{h'}, i \notin PS_{l'} \text{ 혹은 } i \notin S_{h'}, i \in PS_{l'} \text{ 혹은 } i \in S_{h'}, i \notin PS_{l'}$$

$$(i \notin S_{h'} \text{ or } i \notin PS_{l'})$$

$$w_i^{(-k)} = \frac{M_{.l}}{m_{hl}} \frac{1}{M}.$$

4) 경제활동인구조사 2010년 10월 자료를 이용한 분산추정량 비교

현행 통계청에서 사용되고 있는 잭나이프(JK) 분산추정식과 층화집락추출 하에서 산출한 JK 분산추정식(제안된 분산추정식)의 비교를 위하여, 2010년 10월 경제활동인구조사자료를 사용하였다. 해당 자료는 1,810개의 조사구(67,523명)로 구성되어 있다. JK 분산추정식은 JK 반복 가중치군(Jackknife replicate weights set)을 이용하여 계산하며, JK 반복 가중치군은 1,810개의 JK 반복 가중치로 구성되어 있다. 각각의 JK 반복 가중치는 하나의 조사구를 제외한 후 사후층화추정량 계산을 위한 가중치로 이루어진다.

JK 반복 가중치군을 산출하는 방안으로는 총 3가지가 고려되었으며 이는 분산추정식의 계산을 위해 사용되는 최종 가중치와 JK 반복 가중치의 산출 방법에 의하여 결정된다. 첫 번째, JK 반복 가중치군은 A로 표기하며 최종 가중치는 통계청 제공 10월경활설계가중치_작성과정3.xls(표본설계자료)와 ea2010_out.sas7dbat(추계인구자료)을 이용하여 계산되었으며 JK 반복 가중치의 산출 방법은 앞에서 제안된 방법이 사용되었다. 두 번째, JK 반복 가중치군은 B로 최종 가중치는 A와 같은 방법으로 계산되었으며 JK 반복 가중치의 산출 방법은 현행 통계청에서 사용하고 있는 방법을 적용하였다. 마지막, 세 번째 JK 반복 가중치군 C는 ea2010_out.sas7dbat에 만들어져 있는 최종 가중치를 사용하였으며, JK 반복 가중치의 산출방법은 현행 통계청에서 사용하고 있는 방법을 적용하였다.

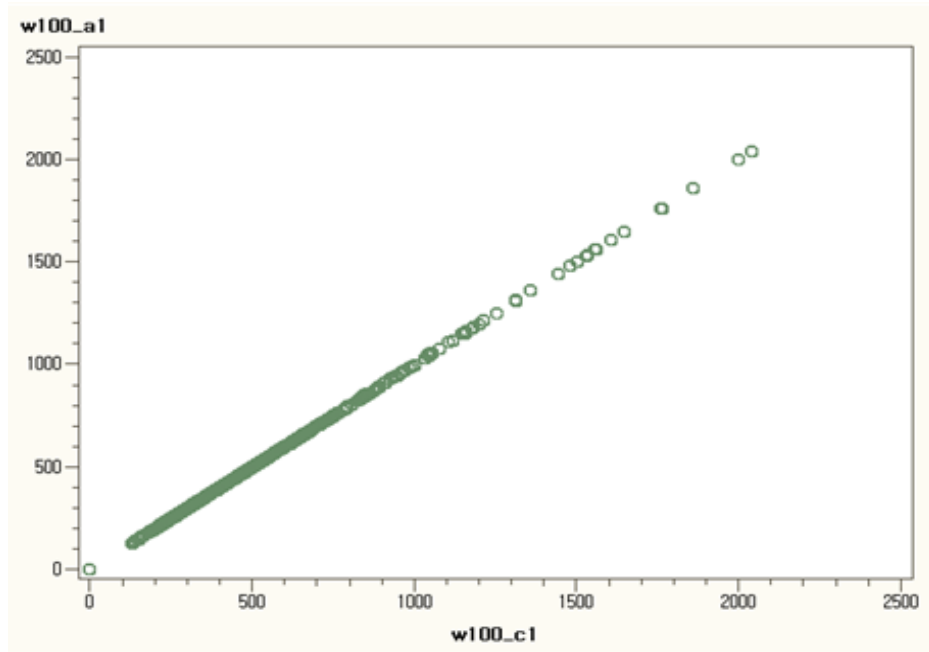
각 방안 별로 산출된 1,810개의 JK 반복 가중치군의 단순 비교를 위해서 특정 조사구(MAJNO3=1115222032)를 제외하고 산출되어진 3가지 방안(A, B, C)의 JK 반복 가중치 w100_a1, w100_b1 그리고 w100_c1을 고려하였다. <표 4-6>은 3가지 방안에 따라서 산출되어진 JK 반복 가중치의 차이에 대한 분포를 나타내고 있다. <표 4-6>에서 same_ac, same_ab 그리고 same_bc는 각각 w100_a1과 w100_c1, w100_a1과 w100_b1 그리고 w100_b1과 w100_c1의 가중치를 비교한 결과이다. w100_a1과 w100_c1 그리고 w100_a1과 w100_b1은 8,903개의 관측치가 서로 다른

가중치를 가지고 있고 w100_b1과 w100_c1은 모든 관측치의 가중치가 같다. 결과적으로 비교를 위해 직접 계산한 최종 가중치와 통계청에서 제공한 최종 가중치는 동일한 반면 통계청에서 사용하는 JK 반복 가중치의 산출 방법(통계청-JK)과 제안된 JK 반복 가중치의 산출 방법(제안된-JK)은 다르다는 것을 알 수 있다. 두 가지의 반복 가중치 산출 방법이 다른 이유는 제안된 JK 반복 가중치는 표본설계 층과 사후 층을 모두 고려해서 JK 반복 가중치를 산출하는 반면에 통계청 JK 반복 가중치는 표본설계 층만을 고려해서 JK 반복 가중치를 산출하기 때문이다. <표 4-6>의 8,903명은 삭제된 조사구에 포함된 가구원(59명)을 제외한 서울지역에 거주하는 가구원 숫자이다.

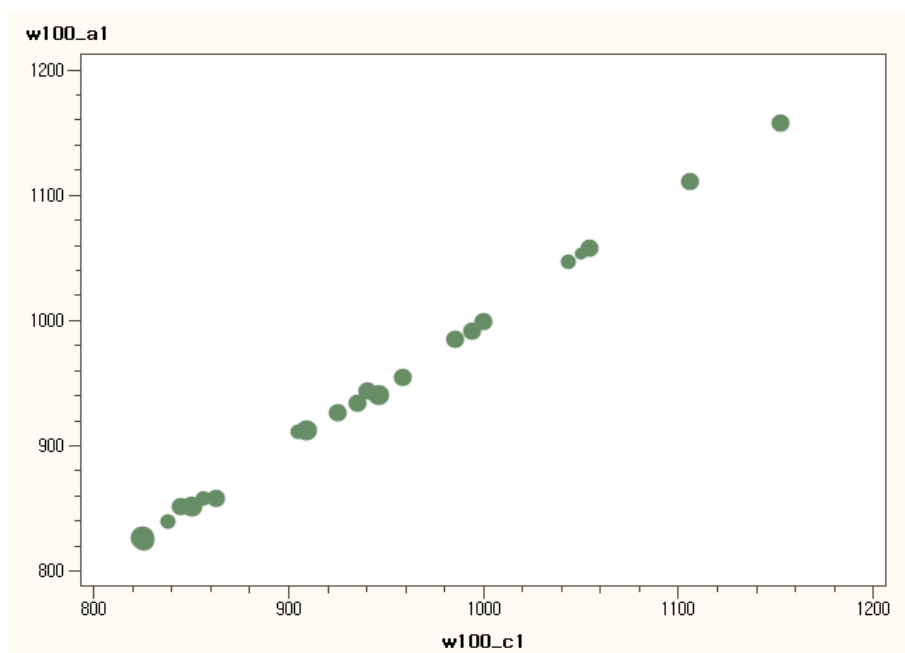
<표 4-6> JK 반복 가중치 비교

구분		빈도	백분율	누적 빈도	누적 백분율
same_ac	diff	8,903	13.19	8,903	13.19
	same	58,620	86.81	67,523	100.00
	합계	67,523	100.00	67,523	100.00
same_ab	diff	8,903	13.19	8,903	13.19
	same	58,620	86.81	67,523	100.00
	합계	67,523	100.00	67,523	100.00
same_bc	same	67,523	100.00	67,523	100.00
	합계	67,523	100.00	67,523	100.00

[그림 4-10]는 모든 관측치 67,523명에 대한 w100_a1과 w100_c1의 산점도이다. 서로 다른 JK 반복 가중치를 갖는 8,903명의 가중치 차이가 크지 않기 때문에 전체적으로는 직선인 것을 확인 할 수 있다. 서로 다른 JK 반복 가중치를 갖는 8,903명의 관측치만을 이용하여 w100_a1과 w100_c1의 <표 4-7>를 작성하고, bubble의 크기를 교차표의 빈도로 하는 BUBBLE Plot을 작성하였다.



[그림 4-10] w100_a1과 w100_c1의 산점도 Plot



[그림 4-11] 가중치가 다른 8,903명의 w100_a1과 w100_c1의 BUBBLE Plot

〈표 4-7〉 가중치가 다른 8,903명의 w100_a1과 w100_c1의 교차표

w100_a1	w100_c1												총합
	824.50	825.53	837.67	844.45	850.27	856.04	862.55	904.63	908.77	924.59	934.76	940.15	
825.15	0	511	0	0	0	0	0	0	0	0	0	0	511
827.11	663	0	0	0	0	0	0	0	0	0	0	0	663
840.55	0	0	256	0	0	0	0	0	0	0	0	0	256
851.86	0	0	0	0	480	0	0	0	0	0	0	0	480
852.56	0	0	0	357	0	0	0	0	0	0	0	0	357
858.70	0	0	0	0	0	267	0	0	0	0	0	0	267
858.80	0	0	0	0	0	0	365	0	0	0	0	0	365
912.13	0	0	0	0	0	0	0	315	0	0	0	0	315
912.52	0	0	0	0	0	0	0	0	470	0	0	0	470
927.14	0	0	0	0	0	0	0	0	0	420	0	0	420
934.77	0	0	0	0	0	0	0	0	0	0	457	0	457
941.64	0	0	0	0	0	0	0	0	0	0	0	0	464
944.53	0	0	0	0	0	0	0	0	0	0	0	442	442
955.67	0	0	0	0	0	0	0	0	0	0	0	0	454
985.88	0	0	0	0	0	0	0	0	0	0	0	0	357
992.01	0	0	0	0	0	0	0	0	0	0	0	0	424
999.74	0	0	0	0	0	0	0	0	0	0	0	0	427
1,047.97	0	0	0	0	0	0	0	0	0	0	0	0	321
1,054.32	0	0	0	0	0	0	0	0	0	0	0	0	219
1,058.79	0	0	0	0	0	0	0	0	0	0	0	0	435
1,111.94	0	0	0	0	0	0	0	0	0	0	0	0	401
1,158.79	0	0	0	0	0	0	0	0	0	0	0	0	398
총합	663	511	256	357	480	267	365	315	470	420	457	442	8,903

〈표 4-7〉 가중치가 다른 8,903명의 w100_a1과 w100_c1의 교차표 (계속)

w100_a1	w100_c1										총합
	945.75	957.73	984.67	994.00	999.42	1,042.80	1,049.34	1,053.73	1,105.77	1,152.27	
825.15	0	0	0	0	0	0	0	0	0	0	511
827.11	0	0	0	0	0	0	0	0	0	0	663
840.55	0	0	0	0	0	0	0	0	0	0	256
851.86	0	0	0	0	0	0	0	0	0	0	480
852.56	0	0	0	0	0	0	0	0	0	0	357
858.70	0	0	0	0	0	0	0	0	0	0	267
858.80	0	0	0	0	0	0	0	0	0	0	365
912.13	0	0	0	0	0	0	0	0	0	0	315
912.52	0	0	0	0	0	0	0	0	0	0	470
927.14	0	0	0	0	0	0	0	0	0	0	420
934.77	0	0	0	0	0	0	0	0	0	0	457
941.64	464	0	0	0	0	0	0	0	0	0	464
944.53	0	0	0	0	0	0	0	0	0	0	442
955.67	0	454	0	0	0	0	0	0	0	0	454
985.88	0	0	357	0	0	0	0	0	0	0	357
992.01	0	0	0	424	0	0	0	0	0	0	424
999.74	0	0	0	0	427	0	0	0	0	0	427
1,047.97	0	0	0	0	0	321	0	0	0	0	321
1,054.32	0	0	0	0	0	0	219	0	0	0	219
1,058.79	0	0	0	0	0	0	0	435	0	0	435
1,111.94	0	0	0	0	0	0	0	0	401	0	401
1,158.79	0	0	0	0	0	0	0	0	0	398	398
총합	464	454	357	424	427	321	219	435	401	398	8,903

특정 조사구를 제외하고 계산되어진 JK 반복 가중치의 단순비교 결과 JK 반복 가중치 w100_b1와 w100_c1은 차이가 없는 것을 알 수 있다. 이는 다른 조사구를 제외하고 계산되어지는 JK 반복 가중치도 같은 결과를 준다. 따라서 JK 반복 가중치군을 이용한 JK 분산추정식의 비교는 방안 A와 C만을 고려하였다. 그리고 선형근사(Taylor approximation) 분산추정식도 같이 비교하였다. 분산추정식의 비교를 위해서는 관심변수인 경제활동인구, 취업인구, 실업인구, 비경제활동인구 그리고 취업시간의 총계를 이용하였다.

<표 4-8> 관심변수의 총계에 대한 표준오차

변수	변수설명	JK 분산추정방안A	JK 분산추정방안C	선형근사(Taylor)
		표준오차	표준오차	표준오차
a	경제활동인구	90,298	210,802	89,548
b	취업인구	90,808	205,592	90,024
c	실업인구	28,846	29,655	28,670
d	비경제활동인구	90,298	170,031	89,548
hsgjno13_3	취업시간	5,088,397	9,819,845	5,043,366

<표 4-8>은 관심변수의 총계 추정량에 대한 표준오차²⁾를 나타내고 있다. JK 분산추정방안 A와 선형근사(Taylor)의 분산추정치는 비슷한 반면, JK 분산추정방안 C의 분산추정치는 다른 두 가지 방안에 비해 매우 큰 것을 알 수 있다. 참고로 통계청에서 제공하고 있는 4개 변수의 JK 표준오차 추정치는 경제활동인구 106,000명, 취업인구 106,000명, 실업인구 29,000명, 비경제활동인구 106,000명이다. 통계청에서 제공하고 있는 표준오차 추정치와 본 연구에서 제공된 결과 즉 방안 C의 결과가 다른 이유는 통계청 방안의 경우, 총합에 대한 사후총화 추정량

$$\hat{t}_{post} = \sum_i w_{i,post} y_i$$

의 분산을 추정하기 위한 공식으로

$$\left(\sum_i w_{i,post} \right)^2 \hat{V} \left(\frac{\sum_i w_{i,post} y_i}{\sum_i w_{i,post}} \right)$$

을 이용하였기 때문이다.

<표 4-9>와 <표 4-10>는 방안 A와 C의 총별 관심변수의 JK 표준오차를 나타내고 있다. 참고로 총별 관심변수의 총계 추정량은 <표 4-11>과 같다.

2) 분산은 수치가 매우 크기 때문에 표준오차로 대치

〈표 4-9〉 총별 JK 분산추정방안 A의 표준오차

표본설계 층	경제활동인구 표준오차	취업인구 표준오차	실업인구 표준오차	비경제활동인구 표준오차	취업시간 표준오차
서울	46,903	46,360	17,166	46,903	2,559,580
부산	22,661	22,344	6,883	22,661	1,200,335
대구	16,710	17,404	5,139	16,710	982,143
인천	17,505	19,449	6,892	17,505	936,295
광주	10,145	10,549	2,716	10,145	559,204
대전	10,116	9,826	3,448	10,116	485,184
울산	8,288	8,056	2,531	8,288	429,689
경기/동부	50,169	49,656	16,231	50,169	2,748,464
경기/읍면	15,906	16,729	4,059	15,906	908,172
강원/동부	7,824	7,864	2,829	7,824	411,421
강원/읍면	7,219	7,266	1,073	7,219	404,850
충북/동부	8,724	8,504	2,501	8,724	424,406
충북/읍면	6,037	6,237	1,022	6,037	412,524
충남/동부	6,435	6,684	2,761	6,435	341,382
충남/읍면	16,657	16,930	3,089	16,657	1,094,482
전북/동부	10,947	10,711	2,516	10,947	551,889
전북/읍면	7,999	7,622	1,926	7,999	345,453
전남/동부	7,195	7,216	1,598	7,195	342,677
전남/읍면	11,887	12,797	2,077	11,887	868,480
경북/동부	9,940	10,137	3,986	9,940	652,557
경북/읍면	15,019	15,727	3,213	15,019	1,053,599
경남/동부	16,439	17,381	5,287	16,439	1,122,270
경남/읍면	16,643	16,825	2,029	16,643	888,671
제주/동부	3,526	3,366	1,214	3,526	194,791
제주/읍면	2,284	2,391	487	2,284	173,586

〈표 4-10〉 총별 JK 분산추정방안 C의 표준오차

표본설계 층	경제활동인구 표준오차	취업인구 표준오차	실업인구 표준오차	비경제활동인구 표준오차	취업시간 표준오차
서울	102,660	100,109	17,509	86,256	4,737,318
부산	50,425	49,278	6,931	41,253	2,287,478
대구	41,247	41,024	5,021	30,341	1,955,924
인천	43,273	42,910	7,028	30,251	2,052,293
광주	25,216	24,883	2,745	16,929	1,157,219
대전	24,112	23,655	3,414	18,849	1,068,278
울산	19,696	19,017	2,615	15,224	892,696
경기/동부	120,888	116,727	16,982	104,546	5,538,609
경기/읍면	39,358	38,759	4,176	29,044	1,856,104
강원/동부	20,103	19,345	2,909	15,485	895,265
강원/읍면	14,329	14,284	1,101	10,852	700,146
충북/동부	22,992	22,527	2,582	16,611	989,872
충북/읍면	19,204	19,006	1,029	10,331	927,496
충남/동부	16,159	15,575	2,870	12,384	721,245
충남/읍면	34,513	33,916	3,232	21,144	1,836,496
전북/동부	21,960	21,264	2,542	21,731	950,058
전북/읍면	19,201	18,379	1,917	12,047	835,089
전남/동부	17,443	17,106	1,658	11,299	742,778
전남/읍면	29,510	29,431	2,167	17,603	1,541,603
경북/동부	28,137	26,873	4,255	21,594	1,291,434
경북/읍면	32,982	32,240	3,704	25,081	1,568,487
경남/동부	46,098	45,406	5,209	33,534	2,284,507
경남/읍면	29,605	29,494	2,049	24,394	1,450,104
제주/동부	11,163	10,986	1,141	7,840	517,942
제주/읍면	4,791	4,855	472	3,760	218,042

〈표 4-11〉 층별 관심변수의 총계 추정

표본설계 층	경제활동인구	취업인구	실업인구	비경제활동인구	취업시간
서울	5,211,203	4,972,420	238,783	3,185,482	231,812,732
부산	1,614,239	1,560,371	53,868	1,285,509	72,772,965
대구	1,209,631	1,168,422	41,209	843,582	54,254,174
인천	1,405,836	1,332,421	73,415	826,708	63,053,777
광주	675,108	660,304	14,804	495,943	30,273,861
대전	735,072	713,011	22,062	500,188	32,987,487
울산	558,194	540,999	17,194	358,361	25,207,387
경기/동부	4,949,001	4,775,260	173,742	3,112,585	222,319,709
경기/읍면	1,044,212	1,018,818	25,393	571,356	47,733,732
강원/동부	411,996	396,988	15,008	302,802	18,352,193
강원/읍면	310,402	307,419	2,983	167,200	13,979,093
충북/동부	435,805	424,629	11,176	307,019	19,342,033
충북/읍면	333,005	329,402	3,603	163,618	15,190,031
충남/동부	366,788	353,677	13,111	236,330	16,816,126
충남/읍면	666,548	651,465	15,082	350,444	31,215,309
전북/동부	533,076	520,572	12,504	419,318	24,058,682
전북/읍면	319,663	314,821	4,843	157,279	13,693,012
전남/동부	347,111	336,022	11,089	232,590	15,240,741
전남/읍면	563,687	557,094	6,593	282,870	25,893,505
경북/동부	688,066	667,158	20,908	410,715	31,401,177
경북/읍면	737,230	722,937	14,293	346,071	34,903,772
경남/동부	958,804	930,383	28,422	646,965	43,961,313
경남/읍면	640,137	632,275	7,862	358,737	29,495,642
제주/동부	201,019	198,070	2,949	109,489	8,949,739
제주/읍면	88,109	87,029	1,081	40,289	3,903,090
총계	25,003,942	24,171,968	831,974	15,711,450	1,126,811,282

5) 모의실험

앞 절에서는 특정 월의 관심변수들의 사후총화추정량을 위한 분산추정량들을 비교하였으나 각 분산추정량들의 평균적인 통계적 성질을 비교하기 위하여 본 절에서는 모의실험을 실행하였다. 비교를 위해 고려된 분산추정량은 JK 추정방안 A, JK 추정방안 C 그리고 선형근사 추정방안이다. 모의실험을 위한 각 구성요소들은 다음과 같다.

모집단: 2010년 10월 경제활동인구조사자료의 서울 지역(230개 조사구)

표본추출 방법: 2단계 집락추출방법(현행 경제활동인구 표본추출방법)

- 1단계: PPS 추출법을 이용한 조사구 추출

- 2단계: 동일확률표본 추출법을 이용한 20개 가구 추출

표본크기: 46개 조사구(모집단의 20%)

조사대상: 추출된 가구에 속한 모든 가구원

사후 층: 22개(연령 11개 수준, 성별 2개 수준)

반복횟수: 5,000번

〈표 4-12〉 총계 추정식 및 분산 추정식의 Monte Carlo Properties(성질)

변수	사후층화추정량의		분산 추정식					
	Monte Carlo 성질		JK 분산추정방안 A		JK 분산추정방안 C		선형근사(Taylor)	
	평균	표준편차	평균	표준편차	평균	표준편차	평균	표준편차
경제활동인구	5,452.91	100.33	114.33	10.96	243.36	27.82	100.88	9.47
취업인구	5,216.09	99.12	112.52	11.48	237.95	26.96	99.35	9.96
실업인구	236.82	33.84	37.51	4.89	37.78	4.83	33.28	4.31
비경제활동인구	3,509.09	100.33	114.33	10.96	209.06	23.28	100.88	9.47
취업시간	242,980.97	5,429.68	6,155.15	572.00	11,270.21	1,263.82	5,439.80	498.34

<표 4-12>을 보면 JK 분산추정방안 A와 선형근사(Taylor)방안에 의하여 산출된 표준오차의 Monte Carlo 평균은 총계에 대한 Monte Carlo 표준편차와 비슷한 반면, JK 분산추정방안 C의 표준오차 추정량의 Monte Carlo 평균은 매우 큰 것을 알 수 있다. 즉 앞 절의 예와 본 절의 모의실험 결과를 통해서 살펴볼 때 현재 사용하고 있는 JK 분산추정방안은 적절하지 않으며 조사구 수만큼의 JK 반복기중치의 산출이 필요한 JK 분산추정량보다는 SAS의 surveyreg 프로시저를 이용한 선형근사 분산 추정량을 사용하는 것이 바람직한 것으로 판단된다.

6) 2상 추출(Two-phase Sampling)하에서의 분산추정

현행 통계청 경제활동인구조사의 표본은 인구주택총조사에서 층화계통추출방법을 이용하여 추출된 10%자료에서 층화집락추출방법을 이용하여 추출된 이상표본(two phase sample)이다. 즉, 인구주택총조사의 10%자료가 1상 표본이 되고 따라서 경제활동인구조사의 표본은 2상 표본이 된다. 1상 표본 추출을 위한 모집단의 표본설계 층은 조사구를 행정구역에 따라 읍면동별로 1차 층화한 후 각 읍면동내에서 조사구의 특성에 따라서 아파트조사구, 보통 및 섬 조사구 그리고 기숙시설 및 특수사회시

설 조사구로 2차 층화하였다. 2상 표본 추출을 위한 1상 표본설계 층은 각각의 도를 읍면부와 동부로 층화하여 18개의 층을 만들고 여기에 7개의 시도를 합하여 총 25개의 층으로 층화 하였다. 통계청에서 가정하고 있는 1상과 2상에서의 표본설계 층은 같지 않다. 따라서 기본적으로 현행 사용하고 있는 분산추정량을 위한 기본적인 가정인 1상 표본추출과 2상 표본추출을 위한 층이 동일하다는 가정은 적절하지 않을 수 있다. 다만 2상 표본 층이 1상 표본 층에 포함됨으로 인하여 서로 다른 표본설계 층의 효과는 그리 크지 않을 수 있다. 본 절에서는 추후 2상 표본 추출법을 통하여 조사가 이루어질 경우, 가능한 사후층화추정량과 이의 분산 추정량이 제시된다.

분산추정방안에 대한 설명을 위해서 모수 추정식 및 분산추정식을 위한 수학적 기호들을 먼저 정의한다. 1상 표본을 위하여 사용된 층을 나타내는 표본설계 층의 기호로는 $h = 1, \dots, H$ 를 사용하고, 2상 표본을 위하여 사용된 층을 나타내는 표본설계 층의 기호로는 $g = 1, \dots, G$ 를 사용한다. 또한 모수를 추정하기 위한 추정식으로는 사후층화추정량을 고려하므로 사후 층을 나타내는 기호로는 $l = 1, \dots, L$ 을 사용한다. 주어진 모집단, 1상 표본 그리고 2상 표본을 각각 U , S_1 그리고 S_2 로 나타낸다. 이상 표본을 이용하여 정의되는 사후층화추정량은 다음과 같은 회귀추정량의 형태로 표현 할 수 있다.

$$\bar{y}_{post} = \bar{y}_{REE} + (\bar{X}_M - \bar{X}_{REE})' \hat{B},$$

여기서

$$\begin{aligned} \bar{y}_{REE} &= \sum_{g=1}^G \sum_{i \in S_2} \left(\frac{\sum_{j \in S_1} w_j a_{jg}}{\sum_{j \in S_2} w_j a_{jg}} \right) w_i a_{ig} y_i \\ &= \sum_{i \in S_2} \alpha_i y_i, \\ \alpha_i &= \sum_{g=1}^G \left(\frac{\sum_{j \in S_1} w_j a_{jg}}{\sum_{j \in S_2} w_j a_{jg}} \right) w_i a_{ig}, \\ a_{ig} &= \begin{cases} 1, & i \in g \\ 0, & i \notin g \end{cases}, \\ \bar{X}_{REE} &= \sum_{i \in S_2} \alpha_i X_i, \\ \bar{X}_M &= \frac{1}{M} \sum_{i=1}^M X_i, \\ X_i &= (\delta_{i1}, \dots, \delta_{iL}), \end{aligned}$$

$$\delta_{il} = \begin{cases} 1, & i \in l \\ 0, & i \notin l \end{cases}$$

$$\hat{B} = \left(\sum_{i \in S_2} \alpha_i X_i' X_i \right)^{-1} \left(\sum_{i \in S_2} \alpha_i X_i' y_i \right),$$

이며 w_i 은 1상 표본의 표본설계 가중치를 그리고 M 은 모집단 경제활동인구 수를 각각 나타낸다. $\bar{y}_{REE}$ 추정량은 Reweighted expansion estimator(REE, Kott & Stukel(1997))로 불리며, Rao & Shao(1992)는 REE 추정량에 대한 반복 분산추정량의 일치성을 보였다.

잭나이프 분산추정량을 구하기 위하여서는 1상 표본의 추출단위인 조사구를 삭제한 후 잭나이프 반복 가중치를 산출하고 이를 이용하여 계산되어진 표본 조사구 수와 동일한 사후층화추정량들의 변동을 계산하여야 한다. 이러한 과정을 통해 정의되는 잭나이프 분산추정량은 다음과 같다.

$$\hat{V}_{JK}(\bar{y}_{post}) = \sum_{k \in S_1} c_k \left(\bar{y}_{post}^{(-k)} - \bar{y}_{post} \right)^2$$

여기서

$$\bar{y}_{post} = \frac{1}{n} \sum_{k=1}^n \bar{y}_{post}^{(-k)},$$

$$\bar{y}_{post}^{(-k)} = \bar{y}_{REE}^{(-k)} + (\bar{X}_M - \bar{X}_{REE}^{(-k)})' \hat{B}^{(-k)},$$

$$\bar{y}_{REE}^{(-k)} = \sum_{i \in S_2} \alpha_i^{(-k)} y_i,$$

$$\alpha_i^{(-k)} = \sum_{g=1}^G \left(\frac{\sum_{j \in S_1} w_j^{(-k)} a_{jg}}{\sum_{j \in S_2} w_j^{(-k)} a_{jg}} \right) w_i^{(-k)} a_{ig},$$

$$w_i^{(-k)} = \begin{cases} 0, & i = k \\ \frac{n_h}{n_h - 1} w_i, & i \neq k \text{ \& } i, k \in S_{1h} \\ w_i, & i \neq k \text{ \& } i \notin S_{1h} \text{ \& } k \in S_{1h} \end{cases},$$

$$\hat{B}^{(-k)} = \left(\sum_{i \in S_2} \alpha_i^{(-k)} X_i' X_i \right)^{-1} \left(\sum_{i \in S_2} \alpha_i^{(-k)} X_i' y_i \right),$$

$$c_k = \frac{n_h - 1}{n_h},$$

이다. 고려된 이상 추출 하에서의 사후층화 추정량과 분산 추정량은 가능한 추정들 중 하나로서 이론적인 내용이 기존 연구자들을 통해 검증된 추정량들이다.

V. 결론

표본설계 관점에서 본 연구를 통해 얻은 결론을 요약하면 다음과 같다. 우선 현행 경제활동인구조사에서 조사구당 20가구를 조사하는 방식이 적절한지 다시 한 번 검토해 볼 필요가 있다. 본 연구결과에 의하면 경제활동인구조사의 경우 개략적으로 조사구당 15가구가 적정 수준인 것으로 분석되었으며, 이에 따라 현행 목표정도를 유지하기 위한 표본크기를 제시하였다. 아울러 현행 36개월에 걸쳐 연동이 이루어지는 대신 30개월을 주기로 연동이 이루어지는 새로운 연동 모형을 제안하였다. 하지만 제안된 연동 모형을 적용하기 위해서는 기존의 방식보다 업무량이 늘어나기 때문에 지방청 실사 담당 인력의 확충이 선행되어야 한다는 점을 반드시 유념해야 할 것이다. 이와 같은 업무량 증가에 따른 문제점은 본 연구결과에 수록되어 있는 지방청 조사직원을 대상으로 한 표적집단면접(FGI) 결과에 잘 요약되어 있다.

아울러 통계청 가구부문 표본설계에서 내재적 충화를 목적으로 적용하는 층 내 조사구 정렬방식과 관련해 기존의 정렬 변수에 추가해 조사구별 30~40대 비율과 30대 이상 대졸자비율 등의 활용이 효율적이라는 결론을 얻을 수 있었다. 한편 현행 경제활동인구조사 응답 패턴 분석결과에 의하면 20~30대 청년층의 경우 장기간에 걸쳐 조사에 참여하는 비율이 상대적으로 다른 연령층에 비해 떨어진다는 것을 확인할 수 있었다. 이와 같이 표본에 포함된 청년층 가구원 변동이 심해지면 결과적으로 안정적인 통계 작성에 영향을 줄 수 있다는 점을 고려해 향후 이런 측면에 대한 보다 심층적인 분석과 함께 관련 무응답률을 줄일 수 있는 실사 관리가 필요할 것으로 보인다.

현행 사용되고 있는 사후층화추정량은 그 효율성, 계산의 용이성을 고려할 때 적절한 추정량으로 판단된다. 가능한 또 다른 추정량으로 각 연동그룹별 추정량에 대한 시계열 모형 적합을 통한 모형 기반의 최량선형불편추정량을 고려할 수 있다. 그

그러나 이의 사용을 위해서는 장기적인 통계청 내의 연구진에 의한 연구가 필요하리라 생각된다. 사후총화추정량의 계속적인 사용을 위해서는 이 후 경제활동인구조사 등을 위해 사용된 표본설계 방법, 즉 이상추출법을 고려한 표본 가중치의 계산 및 이를 바탕으로 한 사후총화추정량을 고려해야 할 것이다.

분산추정량의 산출을 위해서 현재 사용하고 있는 잭나이프 분산추정량은 적절한 반복 가중치가 사용되지 않았으며 그 계산도 복잡하다. 현재 분산추정을 위해 표본 추출과 관련한 통계청의 가정 하에서 통계 이론, 모의실험 그리고 실제 예를 통해서 볼 때 선형 근사 추정량의 사용이 적절하리라 여겨지며 이는 또한 이미 SAS의 회귀 추정량과 이의 분산추정량을 산출하기 위한 프로시저를 사용함으로 그 계산의 복잡성을 피할 수 있으리라 여겨진다. 모수 추정의 경우와 마찬가지로 분산추정에 있어서도 본 연구에서 제시한 이상추출법 하에서의 분산추정량을 참조하여 장기적인 통계청 내부의 연구를 통해 이론적으로 타당하며 그 계산이 용이한 분산추정방안을 고려해야 할 것이다.

■ 참고문헌 ■

- 박유성 · 김기환 《SAS/ETS를 이용한 시계열자료분석 I》 자유아카데미.
- 정미옥 · 김서영 · 정동명 · 정구현. 2008. 《경제활동인구조사의 가중값 조정》.
- 통계청. 2002. “연동표본연구”.
- 통계청. 2003. “가구부문 표본개편보고서”.
- 통계청. 2007. “가구부문 표본개편보고서”.
- 통계청. 2010. “가구부문 경상조사 3차 연동표본 설계보고서”.
- Bell, P. A. and Carolan, A. M. 1998. "TREND ESTIMATION FOR SMALL AREAS from a continuing survey with controlled sample overlap".
- Bell, P. 2001. "Comparison of Alternative Labour Force Survey Estimators" *Survey Methodology*, 27: 53-63.
- Evans, D. L. and Kassinger, T. W. and Cooper, K. B. and Kincannon, C. L. 2004. "Accuracy and Coverage Evaluation of Census 2000: Design and Methodology" U.S. Census Bureau.
- Fuller, W. A. 2001. "A regression composite estimator with application to the Canadian Labour Force Survey" *Survey Methodology*, 27: 45-51.
- Kott, P.S. and Stukel, D.M. 1997. "Can the Jackknife Be Used With a Two-Phase Sample?" *Survey Methodology*, 23, pp. 81-89.
- Rao, J.N.K. and Shao, J. 1992. "Jackknife Variance Estimation with Survey Data Under Hot Deck Imputation" *Biometrika*, 79, pp. 811-822.

부 록

〈부록-표 1〉 (경찰) 경찰인구의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$)별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	1.38	0.024	10	14	16	17	20	22
부산	1.75	0.045	7	10	11	12	14	16
대구	1.95	0.059	6	9	10	11	12	14
인천	1.69	0.044	7	10	12	13	15	16
광주	1.87	0.057	6	9	10	11	13	15
대전	1.75	0.046	7	10	11	12	14	16
울산	1.83	0.049	7	9	10	11	13	15
경기/동부	1.28	0.017	13	18	20	22	25	28
경기/읍면	1.57	0.035	9	12	14	15	17	19
강원/동부	1.49	0.030	10	14	16	17	20	23
강원/읍면	1.68	0.039	8	11	12	13	15	17
충북/동부	1.75	0.046	7	10	11	12	14	16
충북/읍면	1.98	0.056	6	9	10	11	13	14
충남/동부	1.54	0.033	8	12	13	14	17	19
충남/읍면	1.68	0.040	8	11	12	13	15	17
전북/동부	1.25	0.015	17	23	26	29	33	37
전북/읍면	1.83	0.047	8	12	13	14	16	18
전남/동부	1.88	0.052	7	9	10	11	13	15
전남/읍면	2.05	0.062	6	8	9	10	12	13
경북/동부	1.73	0.043	7	10	12	13	15	16
경북/읍면	1.76	0.044	8	11	12	13	15	17
경남/동부	2.26	0.080	5	7	8	9	11	12
경남/읍면	1.73	0.044	8	12	13	15	17	19
제주/동부	1.27	0.016	17	24	26	29	33	37
제주/읍면	1.11	0.005	19	27	30	33	38	42

〈부록-표 2〉 (경찰) 취업인구의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$)별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	1.37	0.023	10	14	16	17	20	22
부산	1.74	0.045	7	10	11	12	14	16
대구	2.01	0.062	6	8	9	10	12	13
인천	1.74	0.047	7	10	11	12	14	16
광주	1.94	0.061	6	9	10	11	12	14
대전	1.80	0.049	7	10	11	12	14	15
울산	1.84	0.050	7	9	10	11	13	15
경기/동부	1.23	0.015	14	19	22	24	27	31
경기/읍면	1.54	0.033	9	12	14	15	18	20
강원/동부	1.53	0.033	10	15	16	18	21	23
강원/읍면	1.70	0.041	7	10	12	13	15	16
충북/동부	1.72	0.044	7	11	12	13	15	17
충북/읍면	1.98	0.056	6	9	10	11	13	14
충남/동부	1.46	0.028	9	13	14	16	18	20
충남/읍면	1.70	0.042	7	11	12	13	15	17
전북/동부	1.26	0.016	14	20	22	24	28	31
전북/읍면	1.82	0.046	8	12	13	14	17	19
전남/동부	1.91	0.054	6	9	10	11	13	14
전남/읍면	2.08	0.063	6	8	9	10	12	13
경북/동부	1.71	0.043	7	10	12	13	15	17
경북/읍면	1.72	0.042	8	11	12	14	16	18
경남/동부	2.26	0.080	5	7	8	9	11	12
경남/읍면	1.68	0.040	9	13	14	16	18	20
제주/동부	1.24	0.014	12	17	19	20	23	26
제주/읍면	1.13	0.007	23	32	36	40	46	51

〈부록-표 3〉 (경활) 비경활인구의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$)별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	2.16	0.073	5	8	9	10	11	12
부산	1.64	0.039	8	11	12	13	15	17
대구	1.79	0.049	7	10	11	12	14	15
인천	1.84	0.054	7	9	10	11	13	15
광주	1.58	0.038	8	11	13	14	16	18
대전	1.53	0.032	8	12	13	15	17	19
울산	1.49	0.029	9	12	14	15	18	20
경기/동부	2.39	0.087	5	7	8	9	10	11
경기/읍면	1.81	0.049	7	10	11	13	15	16
강원/동부	1.69	0.042	8	11	12	13	15	17
강원/읍면	1.53	0.031	9	13	14	15	18	20
충북/동부	1.63	0.038	8	11	13	14	16	18
충북/읍면	1.42	0.024	10	14	15	17	19	22
충남/동부	1.77	0.048	7	10	11	12	14	16
충남/읍면	1.91	0.054	6	9	10	11	13	14
전북/동부	1.46	0.028	9	13	14	16	18	20
전북/읍면	1.92	0.052	6	9	10	11	13	14
전남/동부	1.39	0.023	10	14	16	17	20	22
전남/읍면	1.71	0.042	7	10	11	13	14	16
경북/동부	1.29	0.017	12	16	18	20	23	26
경북/읍면	2.05	0.061	6	8	9	10	12	13
경남/동부	1.58	0.037	8	12	13	14	17	19
경남/읍면	1.63	0.038	8	12	13	14	16	18
제주/동부	1.09	0.005	14	20	23	25	29	32
제주/읍면	1.72	0.038	8	12	13	14	16	18

〈부록-표 4〉 (경찰) 취업시간의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$)별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	1.35	0.025	11	15	17	19	21	24
부산	1.42	0.032	10	15	17	18	21	23
대구	1.73	0.056	7	10	12	13	15	16
인천	1.16	0.012	16	23	26	29	33	37
광주	1.42	0.035	10	14	15	17	19	21
대전	1.11	0.008	17	25	28	30	35	39
울산	1.26	0.018	13	18	20	22	26	29
경기/동부	1.36	0.025	11	15	17	19	22	24
경기/읍면	1.51	0.039	10	15	16	18	21	23
강원/동부	1.26	0.021	11	16	17	19	22	25
강원/읍면	1.90	0.072	6	9	10	11	13	14
충북/동부	1.12	0.009	13	19	21	23	26	29
충북/읍면	2.56	0.121	5	7	8	8	10	11
충남/동부	1.31	0.021	24	34	38	41	48	53
충남/읍면	2.06	0.081	6	9	10	11	12	14
전북/동부	1.52	0.040	10	14	15	17	19	21
전북/읍면	2.73	0.132	5	7	8	8	10	11
전남/동부	1.58	0.044	9	12	14	15	17	19
전남/읍면	3.37	0.179	4	5	6	6	7	8
경북/동부	1.39	0.029	15	21	24	26	30	33
경북/읍면	2.92	0.149	4	6	7	8	9	10
경남/동부	1.29	0.022	14	19	22	24	27	31
경남/읍면	2.28	0.097	6	8	9	10	11	13
제주/동부	1.48	0.032	10	13	15	17	19	21
제주/읍면	1.70	0.045	9	13	14	16	18	20

〈부록-표 5〉 (가계) 소득의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$) 별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	2.24	0.221	3	5	5	6	7	8
부산	1.36	0.060	8	11	12	13	15	17
대구	2.12	0.171	4	5	6	6	7	8
인천	1.54	0.085	6	8	9	10	11	13
광주	2.26	0.216	3	5	5	6	7	7
대전	2.46	0.231	3	5	5	6	7	7
울산	1.88	0.141	7	10	11	12	14	16
경기/동부	2.09	0.168	6	9	10	11	12	14
경기/읍면	1.58	0.098	6	8	9	10	12	13
강원/동부	1.34	0.054	8	11	12	13	15	17
강원/읍면	1.95	0.143	4	6	7	7	9	10
충북/동부	1.92	0.141	5	6	7	8	9	10
충북/읍면	1.67	0.125	5	7	8	8	10	11
충남/동부	2.22	0.200	3	5	6	6	7	8
충남/읍면	2.20	0.188	4	6	6	7	8	9
전북/동부	1.54	0.085	6	8	9	10	11	12
전북/읍면	2.87	0.336	3	4	4	4	5	6
전남/동부	1.97	0.153	5	7	7	8	9	10
전남/읍면	1.93	0.160	4	6	7	7	8	9
경북/동부	2.27	0.190	3	5	5	6	7	8
경북/읍면	2.20	0.214	3	5	5	6	7	8
경남/동부	2.37	0.228	3	5	5	6	6	7
경남/읍면	1.83	0.134	5	7	7	8	9	10
제주/동부	1.72	0.122	5	7	8	9	10	11
제주/읍면	1.33	0.059	11	16	18	19	22	25

〈부록-표 6〉 (가계) 경상소득의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$) 별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	2.35	0.241	3	5	5	6	6	7
부산	1.52	0.087	6	8	9	10	11	12
대구	2.28	0.196	3	5	5	6	7	7
인천	1.81	0.127	5	7	7	8	9	11
광주	2.31	0.224	3	5	5	6	6	7
대전	2.63	0.256	3	4	4	5	6	6
울산	1.97	0.155	4	6	7	8	9	10
경기/동부	2.31	0.203	3	5	5	6	7	7
경기/읍면	1.70	0.118	5	7	8	8	10	11
강원/동부	1.39	0.063	7	10	11	12	14	15
강원/읍면	2.13	0.170	4	5	6	6	7	8
충북/동부	2.12	0.172	4	5	6	6	7	8
충북/읍면	1.71	0.133	5	7	7	8	9	10
충남/동부	2.44	0.238	3	4	5	5	6	7
충남/읍면	2.34	0.210	4	5	6	6	7	8
전북/동부	1.78	0.124	5	6	7	8	9	10
전북/읍면	2.99	0.360	2	3	4	4	5	5
전남/동부	2.02	0.160	5	7	8	8	10	11
전남/읍면	1.95	0.162	4	6	6	7	8	9
경북/동부	2.39	0.208	3	5	5	6	7	7
경북/읍면	2.29	0.230	3	5	5	6	7	7
경남/동부	2.55	0.258	3	4	5	5	6	7
경남/읍면	1.85	0.138	5	6	7	8	9	10
제주/동부	1.94	0.159	4	6	7	7	8	9
제주/읍면	1.23	0.041	14	20	23	25	29	32

〈부록-표 7〉 (가계) 근로소득의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$) 별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	1.74	0.132	5	7	8	8	10	11
부산	1.18	0.030	14	20	22	24	28	31
대구	1.62	0.094	5	7	8	9	10	12
인천	1.59	0.092	5	8	9	10	11	12
광주	1.49	0.084	6	9	10	10	12	13
대전	1.87	0.135	4	6	7	8	9	10
울산	2.07	0.172	4	5	6	7	8	9
경기/동부	1.95	0.148	4	6	6	7	8	9
경기/읍면	1.62	0.104	5	7	8	9	10	12
강원/동부	1.37	0.058	7	10	11	12	14	16
강원/읍면	2.36	0.206	3	5	5	6	6	7
충북/동부	1.71	0.110	5	7	7	8	9	11
충북/읍면	1.69	0.129	5	7	7	8	9	10
충남/동부	1.72	0.118	5	7	7	8	9	11
충남/읍면	2.08	0.170	4	5	6	7	8	9
전북/동부	1.42	0.067	7	10	11	12	14	15
전북/읍면	2.00	0.181	4	5	6	6	8	8
전남/동부	1.54	0.084	6	8	9	10	11	12
전남/읍면	1.87	0.151	4	6	7	7	9	10
경북/동부	2.37	0.205	3	5	5	6	6	7
경북/읍면	2.04	0.186	4	5	6	6	7	8
경남/동부	1.59	0.097	6	9	10	11	12	14
경남/읍면	1.89	0.146	4	6	7	8	9	10
제주/동부	1.81	0.136	4	6	7	8	9	10
제주/읍면	0.77	-0.042	5	7	8	9	10	11

〈부록-표 8〉 (가계) 가계지출의 층별 조사구당 최적 표본 가구수

층	deff	R_u^2	조사비용의 비율($C_{psu}:C_{ssu}$) 별 가구수					
			2:1	4:1	5:1	6:1	8:1	10:1
서울	2.06	0.187	4	5	6	6	7	8
부산	1.52	0.087	7	10	11	12	14	16
대구	2.25	0.192	3	5	6	6	7	8
인천	1.85	0.135	5	6	7	8	9	10
광주	2.03	0.177	4	6	6	7	8	9
대전	2.09	0.171	4	6	6	7	8	9
울산	1.49	0.080	7	9	11	12	13	15
경기/동부	2.06	0.164	4	6	6	7	8	9
경기/읍면	1.42	0.071	8	11	12	13	15	17
강원/동부	1.50	0.079	7	10	11	12	14	15
강원/읍면	1.73	0.110	5	7	8	9	10	11
충북/동부	1.70	0.109	5	7	8	9	11	12
충북/읍면	1.44	0.081	8	11	13	14	16	18
충남/동부	1.89	0.147	4	6	7	7	8	9
충남/읍면	1.62	0.098	8	11	12	13	15	17
전북/동부	1.72	0.113	6	8	9	10	11	12
전북/읍면	2.48	0.268	3	5	5	6	7	7
전남/동부	1.82	0.128	5	7	7	8	9	10
전남/읍면	1.98	0.167	4	6	6	7	8	9
경북/동부	2.14	0.172	4	5	6	6	7	8
경북/읍면	2.06	0.189	4	5	6	7	8	8
경남/동부	2.06	0.176	4	6	7	7	8	9
경남/읍면	1.56	0.091	6	8	9	10	12	13
제주/동부	1.42	0.070	6	9	10	11	12	14
제주/읍면	1.03	0.007	8	11	12	13	15	17